

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:32:13 ; Search time 2266.09 Seconds
(without alignments)
291.578 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagccgttctaagcgccgc 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4538634 seqs, 1887831982 residues

Total number of hits satisfying chosen parameters: 9077268

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

59: gb_est33:*

60: gb_est34:*

61: gb_est35:*

62: gb_est36:*

63: gb_est37:*

64: gb_est38:*

65: em_est27:*

66: em_est28:*

67: em_est29:*

68: em_est30:*

69: gb_est39:*

70: gb_est40:*

71: gb_est41:*

72: gb_est42:*

73: gb_est43:*

74: gb_est44:*

75: em_est31:*

76: em_est32:*

77: em_est33:*

78: em_est34:*

79: gb_gss1:*

80: gb_gss2:*

81: gb_gss3:*

82: gb_gss4:*

83: em_gss1:*

84: em_gss2:*

85: em_gss3:*

86: em_gss4:*

87: gb_gss5:*

88: gb_gss6:*

89: gb_gss7:*

90: gb_gss8:*

91: gb_gss9:*

92: em_gss5:*

93: em_gss6:*

94: em_gss7:*

95: em_gss8:*

96: em_gss9:*

97: em_gss10:*

98: em_gss11:*

99: gb_gss10:*

100: gb_gss11:*

101: em_gss12:*

102: gb_gss12:*

103: gb_gss13:*

104: gb_gss14:*

105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB ID	Description
	Score	Match	Length		
1	34.2	19.5	203	24	H84039
2	33.8	19.3	413	28	AA085404

H84039 yv88e11.s1
AA085404 zn07f02.s

3 33.2 19.0 319 45 A1355469
 4 33.2 19.0 324 34 A501558
 5 33.2 19.0 327 34 A931139
 6 33.2 19.0 339 34 A524847
 7 33.2 19.0 419 24 H77931
 8 33.2 19.0 496 22 R67613
 9 33 33 515 64 A067833
 10 32.2 18.4 371 29 A191759
 11 32.2 18.4 416 28 A088237
 12 32 18.3 751 104 A0550202
 13 31.6 18.1 236 40 A911691
 14 31.6 18.1 251 34 A522480
 15 31.6 18.1 273 45 A1336599
 16 31.6 18.1 316 38 A814534
 17 31.6 18.1 326 47 A1492862
 18 31.6 18.1 334 43 A1190656
 19 31.6 18.1 338 40 A961909
 20 31.6 18.1 344 23 H45269
 21 31.6 18.1 345 43 A1241610
 22 31.6 18.1 385 20 T28751
 23 31.6 18.1 387 45 A1368945
 24 31.6 18.1 387 74 A183573
 25 31.6 18.1 388 47 A1475499
 26 31.6 18.1 403 45 A1350245
 27 31.6 18.1 411 20 T03450
 28 31.6 18.1 442 34 A484227
 29 31.6 18.1 448 40 A961667
 30 31.6 18.1 449 37 A716744
 31 31.6 18.1 449 42 A1086766
 32 31.6 18.1 462 42 A1123673
 33 31.6 18.1 464 45 A1343733
 34 31.6 18.1 467 61 A1865162
 35 31.6 18.1 470 44 A1302423
 36 31.6 18.1 478 37 A706415
 37 31.6 18.1 487 44 A1281253
 38 31.6 18.1 496 44 A1281213
 39 31.6 18.1 504 39 A9831725
 40 31.6 18.1 508 38 A779221
 41 31.6 18.1 512 46 A1440246
 42 31.6 18.1 515 44 A1313149
 43 31.6 18.1 528 40 A948345
 44 31.6 18.1 529 42 A1138558
 45 31.6 18.1 529 64 A078880

ALIGNMENTS

RESULT 1
 H84039 203 bp mRNA EST 13-NOV-1995
 LOCUS yv88a11.s1 Soares melanocyte 2NDHM Homo sapiens cDNA clone
 DEFINITION IMAGE:249836 3', mRNA sequence.
 H84039
 ACCESSION H84039.1 GI:1062710
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 203)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Willson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 High quality sequence stops: 201
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 559 Std Error: 0.00
 Seq primer: Promega -2iml3
 High quality sequence stop: 201.
 Location/Qualifiers
 source 1..203
 /organism="Homo sapiens"
 /db_xref="GDB:3867542"
 /db_xref="taxon:9606"
 /clone="IMAGE:249836"
 /clone_lib="Soares melanocyte 2NDHM"
 /sex="Male"
 /tissue_type="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGAAGTGGGAGCGGCGAGTGTGTTTTTTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes
 (FS374) was kindly provided by Dr. Anthony P. Albino."
 BASE COUNT 49 a 53 c 50 g 44 t 7 Others
 ORIGIN
 Query Match 19.5% Score 34.2 DB 24; Length 203;
 Best Local Similarity 49.4%; Pred.No.2.3; Indels 0; Gaps 0;
 Matches 78; Conservative 0; Mismatches 80;
 Qy 4 ttccaccatgagtcagtatataaagcaaatcttaatttataggtataactgaactag 63
 Db 3 TGCACCAACCTGGCAGGAGGCAACCTGTAGAAAATAAAATTCATGTAAATCCCAAGAT 62
 Qy 64 gagggttcacatggttcgacacacatgctatcctcgtgtaaacaccatcgcggtgttc 123
 Db 63 AGGCGCCCTTGGGTTCATCANCCTTCTCTTCTGTGAGCCCAAGCTCCNCTTTTCN 122
 Qy 124 acggaagtgcagcttcattatcacagagagaacccg 161
 Db 123 NGGGAATGGAGCCCGGAGGAGGCCACAGAGGAAGCAG 160
 RESULT 2
 H84039 413 bp mRNA EST 23-DEC-1997
 LOCUS zn07f02.s1 Stratagene hNT neuron (#937733) Homo sapiens cDNA clone
 DEFINITION IMAGE:546747 3', mRNA sequence.
 H84039
 ACCESSION H84039.1 GI:1627497
 VERSION EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 413)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
 and Warra, M.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On May 5, 1995 this sequence version replaced gi:798595.

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Clone distribution: NCI-CGAP clone d
found through the I.M.A.G.E. Consorti

distribution information c
m/LLNL at:

can be

www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 434 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

Location/Qualifiers
1. .324

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:909465"
/clone_lib="NCI_CGAP_Alvi"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"

/note="Vector: pAMP10; mRNA made from alveolar rhabdomyosarcoma, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."

78 a 79 c 79 g 88 t

BASE COUNT

ORIGIN

Query Match 19.0%; Score 33.2; DB 34; Length 324;
Best Local Similarity 50.6%; Pred. No. 4.5;
Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4 ttccaccatgcatcagatataaaagcaaatctctaaatttatagtgataactgaactag 63

DB 21 TGCACACCATGGCTAAGGTCACAACTTGTAGAAAAATAAAATTCATGTAAATCCCAAGAT 80

QY 64 gaggtgcatcaggttcagacaccatgatactcattcatcgtgtgtaaacattcgcggtgtgc 123

DB 81 AGGGCCCCCTGGTGTTCATCCCTCTCTCTTCTGTGAGCCCAAGTCCTCTCTTTTCT 140

QY 124 acggaagtgcagccttcattatcacagaggaagccg 161

DB 141 TGGGAATGGAGCCCGCAGAGCCACAGAGGAAGCAG 178

RESULT

AA931139

LOCUS

DEFINITION AA931139 327 bp mRNA EST 07-JUL-1998

ACCESSION o070b09.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1571513 3',

VERSION AA931139

KEYWORDS AA931139.1 GI:3085525

SOURCE EST.

ORGANISM human.

REFERENCE Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jan 9, 1998 this sequence version replaced gi:716604.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael

Emmert-Buck, M.D., Ph.D.

cDNA Library prepared by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 458 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 314.

Location/Qualifiers

1. .327

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1571513"

/clone_lib="NCI_CGAP_GC4"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from 3 pooled

germ cell tumors, and was then primed with a Not I -

oligo(dT) primer. Double-stranded cDNA was ligated to Eco

RI adaptors (Pharmacia), digested with Not I and cloned

into the Not I and Eco RI sites of the modified pT73

vector. Library is normalized. Library was constructed by

Bento Soares and M. Fatima Bonaldo."

81 a 81 c 79 g 86 t

Query Match 19.0%; Score 33.2; DB 40; Length 327;

Best Local Similarity 50.6%; Pred. No. 4.5;

Matches 80; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 4 ttccaccatgcatcagatataaaagcaaatctctaaatttatagtgataactgaactag 63

DB 26 TGCACACCATGGCTAAGGTCACAACTTGTAGAAAAATAAAATTCATGTAAATCCCAAGAT 85

QY 64 gaggtgcatcaggttcagacaccatgatactcattcatcgtgtgtaaacattcgcggtgtgc 123

DB 86 AGGGCCCCCTGGTGTTCATCCCTCTCTCTTCTGTGAGCCCAAGTCCTCTCTTTTCT 145

QY 124 acggaagtgcagccttcattatcacagaggaagccg 161

DB 146 TGGGAATGGAGCCCGCAGAGCCACAGAGGAAGCAG 183

RESULT

AA524847

LOCUS

DEFINITION AA524847

ACCESSION AA524847

VERSION AA524847.1

KEYWORDS GI:2265775

SOURCE EST.

ORGANISM human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 339)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On May 5, 1995 this sequence version replaced gi:798189.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,

M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbrrp/image/image.html

Insert Length: 429 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

Location/Qualifiers

1. .339

FEATURES

source

Query Match	Score 32.2	DB 29	Length 371	QY
18.4%	Score 32.2	DB 29	Length 371	4 ttcccaccatggatcagtcagtataaaaagcaaaattcttaatttttaggtataacgaactag 63

64 gaggttgccatggttcagaaccatgtatcatctcatctggtgtaaacaccattcggcggttg 123
|||
86 AGGCCCCCTGGGTNTTCATCCCTCTCTCTTCTGTGAGCCCAAGCTCCTCTCTTTTCT 145
|||

124 acggaagtgagcctgcattatcacagaggaagcgg 161
146 TGCGAATGGAGCCCGCAGAGGCCACAGAGGAAGCAG 183

RESULT 12
0550202

RESULT	12
AQ550202	OOCUS
AQ550202	751 bp DNA GSS 28-MAY-1999
RPCI-11-372I3.TU RPCI-11	Homo sapiens genomic clone RPCI-11-372I3, genomic survey sequence.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 751)

1 (Pages 1 to 727)	
AUTHORS	Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL
COMMENT

Map Building
Unpublished (1997)
On Feb 19, 1999 this sequence version replaced gi.4144141.
Other_GSSs: RPCI-11-37213.TV
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr. Rockville, MD 20850

7/12 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library pBC1-11 For BAC

SOURCE

Sequence information was obtained from the Human BAC library after 10⁶ fold enrichment. Please contact Pieter de Jong for library availability, please contact Pieter de Jong (pieter@jong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html. Seq primer: SP6

CLASS: BAC ends.

LOCATION/QUALIFIERS

1. .751
/organism="Homo sapiens"

```

/db_xref="GDB:7642658"
/db_xref="taxon:9606"
/clone_xref="RPCI-11-37213"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
342 a 104 c 138 g 166 t 1 others
ORIGIN

Query Match 18.3% Score 32; DB 104; Length 751;
Best Local Similarity 59.5%; Prad No 9.8.
```

```

Matches 53; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
best local similarity 39.0%; Fred. NO. 3.0;

```

RESULT	14
AA522480/c	
LOCUS	AA522480/c
DEFINITION	05-AUG-1997 EST IMAGE:982097, mRNA
ACCESSION	AA522480
VERSION	AA522480
KEYWORDS	EST.
SOURCE	AA522480.1 GI:2263192
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1404971. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuquai, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: David B. Krizman, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnl.gov/bbrp/image/image.html

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

```

source
1. .251
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="982097"
/clone_lib="NCI_CGAP_Prl2"
/sex="male"
/tissue_type="metastatic prostate bone lesion"
/lab_host="DH10b"
/note="Vector: pAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NTH"

```

Db	216	TGCAAAACCATGGCTAAGGTCAAAAC	TCTGTAGAAAATAAAATTC	CATGTAATCCCAAAAGAT	157
Qy	64	gaggttgccatggttcagaaaccatgatacattcatcgtggttaaaccattcgcgcggttgtc			123
Db	156	AGGCCCCCTTGGGGTTTCATCCCCCTTCCTCTTCTGTGAGCCCAAGCTCCCTCTCTTTTCCT			97
Qy	124	acggaagtagcccttgctattatacacagagagaaagccg			161
Db	96	TGGGAATGGAGCCCGGCAGAGGCCACAGAAGGAAGCAG			59
RESULT 15					
AI336599					
LOCUS	AI336599	273 bp	nrna	EST	13-FEB-1999
DEFINITION	g062a08.x1 NCI_CGAP_C08	Homo sapiens	cDNA clone	IMAGE:1913078	3'

```

mRNA sequence.
ACCESSION   AI336599
VERSION     AI336599.1  GI:4073526
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 273)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2286499.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1202 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 243.

FEATURES             Location/Qualifiers
     source            1..273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1913078"
                     /clone.lib="NCI CGAP Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT          68 a 70 c 66 g 69 t
ORIGIN
Query Match      18.1%; Score 31.6; DB 45; Length 273;
Best Local Similarity 50.0%; Pred. No. 14;
Matches 79; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY  4  ttccaccatgacgtatataaaagcaaatcttaattatagtgataactgaactag 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   20  TGCATAACCATGGCTAAGTCAAACTTGATAGAAAATAAATTCATGTAATCCCAAGAT 79
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  64  gaggttgccatggttcagaaaccatgatacttcattcgtggttaaacattcgcggtgtc 123
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   80  AGGGGCCCTTGGGGTTTCATCCCTTCTTCTGTGAGCCCAAGCTCTCTCTTTTCT 139
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY  124 acggaagtgccttgcttatacacagaggaagccg 161
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   140 TGGGAATGGAGCCCGCAGAGGCCACAGAGGAGGAGCAG 177
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: April 8, 2000, 01:04:00
Job time: 5507 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:34:45 ; Search time 371.99 Seconds
(without alignments)
-1428.445 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatggtacg.....aagcgttctaagcgccgc 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 821193 seqs, -1518192014 residues
Total number of hits satisfying chosen parameters: 1642386

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_on.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_vi.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48.2	27.5	588	5	A73353	A73353 Sequence 3
2	48.2	27.5	588	5	AR013794	AR013794 Sequence
3	48.2	27.5	588	5	AR047910	AR047910 Sequence
4	48.2	27.5	588	5	I13096	I13096 Sequence 3
5	48.2	27.5	588	5	I25469	I25469 Sequence 3
6	48.2	27.5	591	5	A01422	A01422 mRNA of lam
7	45	25.7	1791	1	CLOFETOX	M12739 C.tetani te
8	45	25.7	1851	5	E01135	E01135 cDNA encodi
9	45	25.7	4338	1	CTTETOXG	X06214 Clostridium
10	45	25.7	4400	1	CTTETANG	X04436 Clostridium
11	41.8	23.9	390	5	E08086	E08086 cDNA encodi
12	41.8	23.9	390	5	E08087	E08087 cDNA encodi
13	41.8	23.9	390	5	E08088	E08088 cDNA encodi
14	41.8	23.9	390	5	E08089	E08089 Modified al
15	41.8	23.9	390	5	E08091	E08091 Modified al
16	41.8	23.9	390	24	E09003	E09003 DNA encodin
17	41.8	23.9	393	5	E06573	E06573 cDNA encodi
18	41.8	23.9	483	34	DEPDER1	D10447 Dermatophag
19	41.8	23.9	485	34	DEPDER3	D10448 Dermatophag
20	41.8	23.9	513	5	AR025156	AR025156 Sequence
21	41.8	23.9	513	5	E03185	E03185 DNA encodin
22	41.8	23.9	516	5	AR025155	AR025155 Sequence
23	41.8	23.9	517	5	E03184	E03184 DNA encodin
24	41.8	23.9	517	34	S70378	S70378 Der f II-gr
25	41.8	23.9	528	34	DEPDER2	D10448 Dermatophag
26	40.2	23.0	491	5	A73357	A73357 Sequence 7
27	40.2	23.0	491	5	AR013796	AR013796 Sequence
28	40.2	23.0	491	5	AR047912	AR047912 Sequence
29	40.2	23.0	491	5	I13098	I13098 Sequence 7
30	40.2	23.0	491	5	I25471	I25471 Sequence 7
31	37	21.1	390	5	E08090	E08090 Modified al
32	33.6	19.2	503	35	AF047613	AF047613 Euroglyph
33	33.6	19.2	505	35	AF047614	AF047614 Euroglyph
C 34	33.4	19.1	2289	34	CELCPR6G	L39939 Caenorhabdi
C 35	33.4	19.1	34514	35	CELC25B8	U41556 Caenorhabdi
C 36	33.4	19.1	65716	11	AC003657	AC003657 Homo sapi
37	33.4	19.1	161814	44	AC016676	AC016676 Homo sapi
C 38	33.4	19.1	172316	44	AC016543	AC016543 Homo sapi
C 39	32.8	18.7	95238	45	AC017105	AC017105 Homo sapi
C 40	32.8	18.7	167561	45	AC016923	AC016923 Homo sapi
C 41	32.4	18.5	3780	34	CEJUNCMECB	L02878 Caenorhabdi
C 42	32.2	18.4	30332	34	CEC43C3	Z47067 Caenorhabdi
C 43	32	18.3	34751	34	CEF26D2	Z81513 Caenorhabdi
C 44	31.8	18.2	1772	7	MTMGCYTBC	X87997 M.galapoda
C 45	31.8	18.2	186598	42	AC006334	AC006334 Homo sapi

ALIGNMENTS

RESULT 1
A73353
LOCUS A73353 588 bp DNA
DEFINITION Sequence 3 from Patent WO9424281.
ACCESSION A73353
VERSION A73353.1 GI:6064126
KEYWORDS

15-OCT-1999

PAT

```

ORGANISM      Unknown:
REFERENCE      1 (bases 1 to 588)
AUTHORS        Garman,R.D., Greenstein,J.L., Kuo,M., Rogers,B.L., Franzen,H.M.,
               Chen,X., Evans,S. and Shaked,Z.
TITLE          T cell epitopes of the major allergens from dermatophagoides (house
               dust mite)
JOURNAL        Patent: US 5820862-A 3 13-OCT-1998;
FEATURES       Location/Qualifiers
               .. 1..588
               /organism="unknown"
BASE COUNT     206 a      105 c      98 g      179 t
ORIGIN
Query Match    27.5%; Score 48.2; DB 5; Length 588;
Best Local Similarity 94.3%; Pred. No. 0.00031;
Matches 50; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 61 taggaggttgccattggttcagaaccattgatcatcttcctcggttaaacattc 113
    |||
Db 172 TACCAGGATGCCAGGTTCAGAACCATGTATCATTCATCGTGGTAAACCATTC 224
    |||

```

[illegible]

REFERENCE	1 (bases 1 to 588)		
AUTHORS	Thomas, W. R. and Chua, K.		
TITLE	Cloning and sequencing of allergens of dermatophagoides (house dust mite)		
JOURNAL	Patent: US 5433948-A 3 18-JUL-1995;		
FEATURES	Location/Qualifiers		
source	1. 588		
	/organism="unknown"		
BASE COUNT	206 a 105 c 98 g		179 t
ORIGIN			

	Query Match	27.5%	Score 48.2	DB 5	Length 588
	Best Local Similarity	94.3%	Pred. No. 0.00031		
	Matches 50	Conservative 0	Mismatches 3	Indels 0	Gaps 0
Qy	61	tagaggttgccatgggttcagagaccatgtatcattcatcggtgtaaacattc	113		
Db	172	TACCAGAGATGCCATGGTTTCAGAACCATGTATCATTCGTGGTAAACCATTC	224		

REFERENCE
AUTHORS 1 (bases 1 to 588)
TITLE Thomas, W.R. and Chua, K.
JOURNAL Cloning and sequencing of allergens of dermatophagoides (house dust
FEATURES mite)
source Patent: US 5552142-A 3 03-SEP-1996;
 Location/Qualifiers
 1. .588
 /organism="unknown"

BASE COUNT	206 a	105 c	98 g	179 t
ORIGIN				
Query Match	27.5%;	Score 48.2;	DB 5;	Length 588;
Best Local Similarity	94.3%;	Pred. No. 0.00031;		
Matches	50;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	61	taggaggtgccatggttcagaaacattatcatcattcgtggtgtaaacattc	113	
Db	172	TACCAGGATCCATGGTTCAGAACCATGTATCATTCATCGTGGTAACCATTC	224	
RESULT	6			
LOCUS	A01422	591 bp	mRNA	PAT
DEFINITION	mRNA of lambda gt11 pII(Cl).			
ACCESSION	A01422			
VERSION	A01422.1	GI:344350		
KEYWORDS	synthetic construct.			
SOURCE	synthetic construct.			
ORGANISM	artificial sequence.			
REFERENCE	1 (bases 1 to 591)			
AUTHORS	CLONING OF MITE ALLERGENS			
TITLE	Patent: WO 8810297-A 3 29-DEC-1988;			
JOURNAL	Location/Qualifiers			
FEATURES	source			
	1..591	/organism="synthetic construct"		
		/db_xref="taxon:32630"		
	72..509	/codon_start=1		
		/transl_table=11		
		/product="Der p II"		
		/protein_id="CA00161.1"		
		/db_xref="GI:344351"		
		/translation="MYKILCLSLVAARQDVVDKCANHEIKKVLVPGCHGSEPCI		
		IHRGKPFQLEAVFEANQNTAKIEIKASIDGLEVDVPGIDPNACHYMKCPVLVKGOY		
		DIKYTWNVKPIAKPSENVVTVKVMGDDGVLCALAIATHAKIRD"		
mat_peptide	120..506			
BASE COUNT	206 a	107 c	98 g	180 t
ORIGIN				
Query Match	27.5%;	Score 48.2;	DB 5;	Length 591;
Best Local Similarity	94.3%;	Pred. No. 0.00031;		
Matches	50;	Conservative	0;	Mismatches 3; Indels 0; Gaps 0;
QY	61	taggaggtgccatggttcagaaacattatcatcattcgtggtgtaaacattc	113	
Db	172	TACCAGGATCCATGGTTCAGAACCATGTATCATTCATCGTGGTAACCATTC	224	
RESULT	7			
LOCUS	CLOTETTOX	1791 bp	DNA	BCT
DEFINITION	C.tetani tetanus toxin gene.			
ACCESSION	M12739			
VERSION	M12739.1	GI:144920		
KEYWORDS	tetanus toxin.			
SOURCE	C.tetani (strain CN3911, derivative of Harvard strain) DNA, clones pTet[1.8].			
ORGANISM	Clostridium tetani			
REFERENCE	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.			
AUTHORS	1 (bases 1 to 1791)			
TITLE	Fairweather,N.F., Lyness,V.A., Pickard,D.J., Allen,G.P. and Thomson,R.O.			
JOURNAL	Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli			
	J. Bacteriol. 165, 21-27 (1986)			


```

FT CDS >1..1722 /product='tetanustoxin'
FT mat_peptide >1..1719
FT 3'UTR /product='tetanustoxin'
FT mRNA 1723..1851
FT Location/Qualifiers
  1..1851
    /organism="Clostridium tetani"
    /db_xref="taxon:1513"
BASE COUNT 753 a 204 c 279 g 615 t
ORIGIN

Query Match 25.7%; Score 45; DB 5; Length 1851;
Best Local Similarity 100.0%; Pred. No. 0.0027;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 cagtatataaaagcaaatctctaaatttaggtataactgaacta 62
|||||
Db 262 CAGTATATAAAGCAAAATCTAAATTTAGGTATAACTGAAC TA 306

RESULT 9
CITEFOXG
LOCUS CTTETOXG 4338 bp DNA BCT 12-SEP-1993
DEFINITION Clostridium tetani gene for tetanus toxin.
ACCESSION X06214
VERSION X06214.1 GI:40773
KEYWORDS tetanus toxin.
SOURCE Clostridium tetani.
ORGANISM Clostridium tetani.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4338)
AUTHORS Fairweather, N.F. and Lyness, V.A.
TITLE The complete nucleotide sequence of tetanus toxin
JOURNAL Nucleic Acids Res. 14 (19), 7809-7812 (1986)
MEDLINE 87040747
COMMENT Data kindly reviewed (25-Nov-1988) by Lyness V.A.
FEATURES
  source
    1..4338
    /organism="Clostridium tetani"
    /strain="CN3911"
    /db_xref="taxon:1513"
    /clone="PAT153"
  322..4269
    /note="tetanus toxin (AA 1-1315)"
    /codon_start=1
    /transl_table=11
    /protein_id="CAA29564.1"
    /db_xref="GI:40774"
  CDS
    322..4269
    /note="tetanus toxin (AA 1-1315)"
    /translation="MPITNNFRYSDPVNNTIIMPEPPYCKGLDIYKAFKITDRIW
    IPERYEGTKPEDPNPSSLEGASEYDNPYLRDSDKDRFLQTMVKLFNRKNNV
    AGEALLDKIINAPYLGNISYLLDFDFTNSVSNLLEQDPGSGATTKSAMLTNLIIF
    GPGPVLNKNEVRGIVLRVDKNYPFCRDGFGSIMQAFCEPVYFDNVIENITSLTI
    GSKYFQDPALLMHELHIVLHGLYGMQVSSHEIIPSKQEIYMQHTYPISAELFTFG
    GDANLISIDIKNDLYEKTNDYKAANKLSQVTSNDPNIDIDSYKIOOQYQFDK
    DSNQGYIVNEKFOQLYNSIMYGFTEIELGKFNKTRLSYFSMNDVPKIPNLDDT
    IYNDTEGNIESKDLKSEYKQNMVRNTAFNRVDSGLVSKLIGLCKKIIPTNIRE
    NLYNTASIDLGELCIKINEDLTFIAEKNSRSEEPFQDEIVSYNTKKNLNFNS
    LDKLIVYDNLQSLTLPNDRTPTKGIPIAPYKSNAASTIEIHNIDNTIYQYLYA
    QKSPITLARIWTNSVDALLINVTGIYFVSIVSKVQAGQILFLQVRDIIIDFT
    NESSQKTTIDKISDVSTIVPIGPNALIVKQYEGNFGALETTCVLLLEYIPETL
    PVIALASSTQKREKIYKIDNPLEKRYEKWIEYLRKAKWLGTVNTQPKRSYQ
    MYRSLEYQDAIKKIDYEYKIYSGPDKQIADENLNKLEKANKAMINIFMR
    ESSRLEYQDAIKKIDYEYKIYSGPDKQIADENLNKLEKANKAMINIFMR
    TPISYKSLNLDCHVDNEDIDVILKSTILNLDINDIISDISGNSSVIYPAQL
    VPGINKAHLNVSSESVIVHKAMDIEYNMFNTVSVLRVPKVSASHLELEGQTN
    EYIISLSSKSLSTIGSGVSLKGNLNLITLWLSKAGVEYRQITFDLPKFNAYLANK
    WVFITINDRLSSANLYNGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDK
    FRIFCKALNPKEIEKLYTSITFLRDFWGNPLRYDTEYLYIPVASSSKVOLKNIT
    DYMILTNPASTNGKLNYYRRLYNGLKFIKRYTPNNEIDSVKSGDFIKLYSVNN
    NEHIVGPKDGNAFNDRILRILRYGYNAPGIPLYKKRKAVALKRLDKLTKSYQLKLYDDKN
    ASHGLVGTGHNQIGNDRDILNAGNWFNHLKDKILGCDWTFVPTDEGWIND"
BASE COUNT 1783 a 522 c 640 g 1393 t
ORIGIN

Query Match 25.7%; Score 45; DB 1; Length 4338;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 cagtatataaaagcaaatctctaaatttaggtataactgaacta 62
|||||
Db 2809 CAGTATATAAAGCAAAATCTAAATTTAGGTATAACTGAAC TA 2853

RESULT 10
CITEFANG
LOCUS CTTETANG 4400 bp DNA BCT 12-JUL-1995
DEFINITION Clostridium tetani gene for tetanus toxin.
ACCESSION X04436
VERSION X04436.1 GI:40769
KEYWORDS neurotoxin; plasmid; tetanus toxin; toxin.
SOURCE Clostridium tetani.
ORGANISM Clostridium tetani.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4400)
AUTHORS Eisel, U., Jarausch, W., Goretzki, K., Henschen, A., Engels, J.,
Weller, U., Hudel, M., Habermann, E. and Niemann, H.
TITLE Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins
JOURNAL EMBO J. 5 (10), 2495-2502 (1986)
MEDLINE 87053814
COMMENT See also Fairweather et al. (1986) J.Bacteriol. 165, 21-27. Data
kindly reviewed (12-MAY-1987) by U.Eisel.
FEATURES
  source
    1..4400
    /organism="Clostridium tetani"
    /db_xref="taxon:1513"
    <1..50
    /note="open reading frame (16AA) (1 is 2nd base in codon)"
    /codon_start=3
    /transl_table=11
    /protein_id="CAA28032.1"
    /db_xref="GI:809667"
    /db_xref="SPTREMBL:Q46456"
    /translation="ALKKLEPIVKNLNI"
    37..250
    /note="region of inverted repeats"
  267..272
    /note="rRNA binding site"
  281..1651
    /product="tetanus toxin precursor (AA 1-1315)"
    /protein_id="CAA28033.1"
    /db_xref="GI:40770"
  281..4228
    /codon_start=1
    /transl_table=11
    /product="tetanus toxin precursor (AA 1-1315)"
    /protein_id="CAA28033.1"
    /db_xref="GI:40770"
  /db_xref="SWISS-PROT:P04958"
  /translation="MPITNNFRYSDPVNNTIIMPEPPYCKGLDIYKAFKITDRIW
  IPERYEGTKPEDPNPSSLEGASEYDNPYLRDSDKDRFLQTMVKLFNRKNNV
  AGEALLDKIINAPYLGNISYLLDFDFTNSVSNLLEQDPGSGATTKSAMLTNLIIF
  GPGPVLNKNEVRGIVLRVDKNYPFCRDGFGSIMQAFCEPVYFDNVIENITSLTI
  GSKYFQDPALLMHELHIVLHGLYGMQVSSHEIIPSKQEIYMQHTYPISAELFTFG
  GDANLISIDIKNDLYEKTNDYKAANKLSQVTSNDPNIDIDSYKIOOQYQFDK
  DSNQGYIVNEKFOQLYNSIMYGFTEIELGKFNKTRLSYFSMNDVPKIPNLDDT
  IYNDTEGNIESKDLKSEYKQNMVRNTAFNRVDSGLVSKLIGLCKKIIPTNIRE
  NLYNTASIDLGELCIKINEDLTFIAEKNSRSEEPFQDEIVSYNTKKNLNFNS
  LDKLIVYDNLQSLTLPNDRTPTKGIPIAPYKSNAASTIEIHNIDNTIYQYLYA
  QKSPITLARIWTNSVDALLINVTGIYFVSIVSKVQAGQILFLQVRDIIIDFT
  NESSQKTTIDKISDVSTIVPIGPNALIVKQYEGNFGALETTCVLLLEYIPETL
  PVIALASSTQKREKIYKIDNPLEKRYEKWIEYLRKAKWLGTVNTQPKRSYQ
  MYRSLEYQDAIKKIDYEYKIYSGPDKQIADENLNKLEKANKAMINIFMR
  ESSRLEYQDAIKKIDYEYKIYSGPDKQIADENLNKLEKANKAMINIFMR
  TPISYKSLNLDCHVDNEDIDVILKSTILNLDINDIISDISGNSSVIYPAQL
  VPGINKAHLNVSSESVIVHKAMDIEYNMFNTVSVLRVPKVSASHLELEGQTN
  EYIISLSSKSLSTIGSGVSLKGNLNLITLWLSKAGVEYRQITFDLPKFNAYLANK
  WVFITINDRLSSANLYNGVLMGSAEITGLGAIREDNNITLKLDRCNNNNQYVSIDK
  FRIFCKALNPKEIEKLYTSITFLRDFWGNPLRYDTEYLYIPVASSSKVOLKNIT
  
```

PVTAALSIASSTOKELIKITIDNLEKRYEKWIEYKLVKAKWLGTVNTQFORSVQ
 MRSLEYQVDAKKIDYKIKYISGPDKEQIADEINLNKLEKANKAMININIFMR
 ESRSLVNMINEAKKQLLEFDTQSKNLMQYKANSKPIGTEKLESKINKVFS
 TRPFSYKNDLVNDEEDIDVILKSTILNDINNDIISDGSNSVITYPDQOL
 VPGINGKALHLVNNESSEVIVHKAMDIENMNFNTVFWLRVPKVSAHLYOYGN
 EYXISSMKKHSLISGSGSVSLKGNLWTLKDSAGEVRIQITFRLDPKFNAYLANK
 WYFITNDRLSANLYINGLMSAEITGLGAIREDNNTILKDRCNNNNQVYSIDK
 FRIFCALNPKEIKYISVLSITFLDFWGNELRIDETIYLIPVASSSKDVOLKNT
 DMYLTNPASYNGKLNIIYRRLYNGLKFIKRYTPNNEIDSVFKSGDFIKLYSVNN
 NHHVGYPDGNFNLDRILRYGYNAPGIPLYKMEAVKRLDKTYSVOLKLYDDKN
 ASLGLVGTINGNDPNRDLIASNWFNHLKDKILGCDWYFVPTDEGTWND"
 1652. .4225
 mat_peptide /products=tetanus toxin H-chain (AA 458-1315)"
 misc_feature /note=region of inverted repeats"
 468. .4394
 BASE COUNT 1808 a 532 c 644 g 1416 t
 ORIGIN

Query Match 25.7%; Score 45; DB 1; Length 4400;
 Best Local Similarity 100.0%; Pred. No. 0.0028;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 cagtatataaaccaattctaaatttataggtataactgaacta 62
 |||||
 Db 2768 CAGTATATAAAGCAATCTAAATTTATAGGTATACTGAACTA 2812

RESULT 11
 E08086 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 E08086
 ACCESSION E08086.1 GI:2176211
 VERSION JP 1994253851-A/10.
 KEYWORDS Dermatophagoides farinae.
 SOURCE Dermatophagoides farinae.
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcotiformes; Astigmata; Pyroglyphoidea;
 Pyroglyphidae; Dermatophagoides.
 REFERENCE 1 (bases 1 to 390)
 AUTHORS Takai,T., Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
 TITLE MODIFIED MAJOR MITE ALLERGEN AND ITS PRODUCTION
 JOURNAL Patent: JP 1994253851-A 10 13-SEP-1994;
 ASahi BREWERIES LTD, TORII YAKUHIIN KK, NIKKA UTSUKII KK
 COMMENT OS Dermatophagoides farinae (acarid)
 PN JP 1994253851-A/10
 PD 13-SEP-1994
 PF 04-MAR-1993 JP 1993139793
 PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA
 HIROSHI,
 PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
 (C12P21/02);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT source 1. .390
 FT /organism='Dermatophagoides farinae'

BASE COUNT 125 a 78 c 76 g 111 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaaacattatcatcattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATGGTTCTGATCCATCATCATCCATCGTGTGTAACCATTC 105

RESULT 13
 E08088 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 E08088
 ACCESSION E08088.1 GI:2176213
 VERSION JP 1994253851-A/12.
 KEYWORDS Dermatophagoides farinae.
 SOURCE Dermatophagoides farinae.

BASE COUNT 125 a 78 c 76 g 111 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaaacattatcatcattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATGGTTCTGATCCATCATCATCCATCGTGTGTAACCATTC 105

RESULT 12
 E08087 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 E08087
 ACCESSION E08087.1 GI:2176212
 VERSION JP 1994253851-A/11.
 KEYWORDS Dermatophagoides farinae.
 SOURCE Dermatophagoides farinae.
 ORGANISM
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Acariformes; Sarcotiformes; Astigmata; Pyroglyphoidea;
 Pyroglyphidae; Dermatophagoides.
 REFERENCE 1 (bases 1 to 390)
 AUTHORS Takai,T., Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
 TITLE MODIFIED MAJOR MITE ALLERGEN AND ITS PRODUCTION
 JOURNAL Patent: JP 1994253851-A 11 13-SEP-1994;
 ASahi BREWERIES LTD, TORII YAKUHIIN KK, NIKKA UTSUKII KK
 COMMENT OS Dermatophagoides farinae (acarid)
 PN JP 1994253851-A/11
 PD 13-SEP-1994
 PF 04-MAR-1993 JP 1993139793
 PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA
 HIROSHI,
 PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
 (C12P21/02);
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC Feature is identified by experimental;
 FH Key Location/Qualifiers
 FT source 1. .390
 FT /organism='Dermatophagoides farinae'

BASE COUNT 126 a 77 c 75 g 112 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaaacattatcatcattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATGGTTCTGATCCATCATCATCCATCGTGTGTAACCATTC 105

RESULT 13
 E08088 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 E08088
 ACCESSION E08088.1 GI:2176213
 VERSION JP 1994253851-A/12.
 KEYWORDS Dermatophagoides farinae.
 SOURCE Dermatophagoides farinae.

BASE COUNT 126 a 77 c 75 g 112 t
 ORIGIN

Query Match 23.9%; Score 41.8; DB 5; Length 390;
 Best Local Similarity 86.8%; Pred. No. 0.023;
 Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttgccatggttcagaaacattatcatcattcgttgtaaacattc 113
 |||||
 Db 53 TCGATGGTTCATGGTTCTGATCCATCATCATCCATCGTGTGTAACCATTC 105

RESULT 13
 E08088 390 bp RNA PAT 29-SEP-1997
 LOCUS cDNA encoding major allergens of Dermatophagoides farinae and their
 DEFINITION modified peptides.
 E08088
 ACCESSION E08088.1 GI:2176213
 VERSION JP 1994253851-A/12.
 KEYWORDS Dermatophagoides farinae.
 SOURCE Dermatophagoides farinae.

```

FT      source      1. 390
FT      /organism='Artificial sequences'
FEATURES
source      Location/Qualifiers
1. 390
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      127 a 79 c 76 g 108 t
ORIGIN

Query Match      23.9%; Score 41.8; DB 5; Length 390;
Best Local Similarity 86.8%; Pred. No. 0.023;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttccatggttcagaaccatgtatcattcatcgtggtgaaccattc 113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 TCGATGTTGCCATGGTTCGTATCCATGCATCCATCCATCGTGGTAAACCATTC 105

RESULT 15
E08091      E08091      390 bp      DNA      PAT      29-SEP-1997
LOCUS      Modified allergen DNA of Dermatophagoides farinae.
DEFINITION
ACCESSION      E08091
VERSION      E08091.1 GI:2176216
KEYWORDS      JP 1994253851-A/15.
SOURCE      unidentified.
ORGANISM      unidentified
unclassified.
REFERENCE      1 (bases 1 to 390)
AUTHORS      Takai,T.,Yuki,T., Okumura,Y., Yamakawa,H., Ando,T. and Hirai,M.
TITLE      MODIFIED MAJOR MITE ALLERGEN AND LTS PRODUCTION
JOURNAL      Patent-JP 1994253851-A 15 13-SEP-1994;
ASahi Breweries Ltd, Torii Yakuhin KK, Nikka Utsukii KK
COMMENT      OS None
OC Artificial sequences.
PN JP 1994253851-A/15
PD 13-SEP-1994
PF 04-MAR-1993 JP 1993139793
PI TAKAI TOSHIRO, YUKI TOSHIFUMI, OKUMURA YASUSHI, PI YAMAKAWA HIROSHI,
ANDO TORU, HIRAI MITSUO
PC C12N15/30,A61K39/35,C12N1/21,C12P21/02,(C12N1/21,C12R1:19), PC
(C12P21/02,
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
FH Key
PH Location/Qualifiers
FT      source      1. 390
FT      /organism='Artificial sequences'
FEATURES
source      Location/Qualifiers
1. 390
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT      127 a 78 c 76 g 109 t
ORIGIN

Query Match      23.9%; Score 41.8; DB 5; Length 390;
Best Local Similarity 86.8%; Pred. No. 0.023;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 61 taggaggttccatggttcagaaccatgtatcattcatcgtggtgaaccattc 113
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 TCGATGTTGCCATGGTTCGTATCCATGCATCCATCCATCGTGGTAAACCATTC 105

Search completed: April 8, 2000, 02:40:03
Job time: 11118 sec

```

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:52:11 2000; MasPar time 27.56 Seconds
Tabular output not generated. 15.169 Million cell updates/sec

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 26.377; Variance 28.586; scale 0.923

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	106	100.0	146	1	DER2_DERPT MITE ALLERGEN DER P 2	1.34e+14
2	85	80.2	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	2.19e+08
3	55	51.9	548	1	YDID_ECOLI HYPOTHETICAL 60.8 KD P	1.13e+00
4	55	51.9	912	1	NIA2_HORVU NITRATE REDUCTASE (EC	1.13e+00
5	55	51.9	915	1	NIA1_HORVU NITRATE REDUCTASE (EC	1.13e+00
6	54	50.9	257	1	YJE3_YEAST HYPOTHETICAL 29.2 KD P	1.89e+00
7	54	50.9	396	1	AMOA_AERHY PUTATIVE ISOCHROMATASE	1.89e+00
8	54	50.9	397	1	YBP2_YEAST HYPOTHETICAL 45.5 KD P	1.89e+00
9	54	50.9	787	1	OBP_HSV7J REPLICATION ORIGIN BIN	1.89e+00
10	54	50.9	1184	1	DP3A_MYCTU PROBABLE DNA POLYMERASE	1.89e+00
11	53	50.0	278	1	NIFH_METVO NITROGENASE IRON PROTE	3.12e+00
12	53	50.0	279	1	NIFH_METJA NITROGENASE IRON PROTE	3.12e+00
13	53	50.0	424	1	OAT_YEAST ORNITHINE AMINOTRANSFER	3.12e+00
14	53	50.0	631	1	ETED_YEAST PROBABLE ELECTRON TRAN	3.12e+00
15	52	49.1	188	1	LCAT_PIG PHOSPHATIDYLCHOLINE-ST	5.10e+00
16	52	49.1	287	1	TYS1_MYCPN THYMIDYLATE SYNTHASE (5.10e+00
17	52	49.1	438	1	LCAT_MOUSE PHOSPHATIDYLCHOLINE-ST	5.10e+00
18	52	49.1	440	1	LCAT_RABIT PHOSPHATIDYLCHOLINE-ST	5.10e+00
19	52	49.1	440	1	LCAT_PAPAN PHOSPHATIDYLCHOLINE-ST	5.10e+00
20	52	49.1	440	1	LCAT_HUMAN PHOSPHATIDYLCHOLINE-ST	5.10e+00
21	51	48.1	286	1	CBP_K_PEA SERINE CARBOXYPEPTIDAS	8.30e+00
22	51	48.1	309	1	YQGH_BACSU PROBABLE ABC TRANSPORT	8.30e+00
23	51	48.1	332	1	COBS_PSEDE COBS PROTEIN.	8.30e+00

24	51	48.1	440	1	LCAT_RAT PHOSPHATIDYLCHOLINE-ST	8.30e+00
25	51	48.1	655	1	YQGH_ECOLI HYPOTHETICAL 70.0 KD P	8.30e+00
26	51	48.1	1176	1	CARB_TRICU CARBAMOYL-PHOSPHATE SY	8.30e+00
27	51	48.1	1607	1	LMG1_MOUSE LAMININ-GAMMA-1 CHAIN	8.30e+00
28	51	48.1	2499	1	MPRI_BOVIN CATION-INDEPENDENT MAN	8.30e+00
29	50	47.2	325	1	SVK3_ECOLI HYPOTHETICAL LYSYL-TRN	1.34e+01
30	50	47.2	329	1	GCC_RAT IG GAMMA-2C CHAIN C RE	1.34e+01
31	50	47.2	355	1	FK38_HUMAN -38 KD FK-506 BINDING P	1.34e+01
32	50	47.2	370	1	PSPB_RABIT PULMONARY SURFACTANT-A	1.34e+01
33	50	47.2	373	1	Y007_METJA HYPOTHETICAL PROTEIN M	1.34e+01
34	50	47.2	387	1	YB96_YEAST HYPOTHETICAL 43.3 KD P	1.34e+01
35	50	47.2	432	1	Y175_PYRHO HYPOTHETICAL PROTEIN P	1.34e+01
36	50	47.2	483	1	TR2_HUMAN ORPHAN NUCLEAR RECEPTO	1.34e+01
37	50	47.2	588	1	RTG2_YEAST RETROGRADE REGULATION	1.34e+01
38	50	47.2	778	1	TAST_HUMAN TASTIN (TROPHEININ-ASSI	1.34e+01
39	49	46.2	209	1	GTP_MESAU GLUTATHIONE S-TRANSFER	2.14e+01
40	49	46.2	209	1	GTP1_MOUSE GLUTATHIONE S-TRANSFER	2.14e+01
41	49	46.2	222	1	R11C_TOBAC RAS-RELATED PROTEIN RA	2.14e+01
42	49	46.2	284	1	NIF1_METTL NITROGENASE IRON PROTE	2.14e+01
43	49	46.2	292	1	NIF2_METTL NITROGENASE IRON PROTE	2.14e+01
44	49	46.2	441	1	YLIG_ECOLI HYPOTHETICAL 49.6 KD P	2.14e+01
45	49	46.2	1609	1	LMG1_HUMAN LAMININ-GAMMA-1 CHAIN	2.14e+01

ALIGNMENTS

RESULT 1
ID DER2_DERPT STANDARD; PRT; 146 AA.
AC P49278;
DT 01-FEB-1996 (Rel. 33; Created)
DT 01-FEB-1996 (Rel. 33; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
GN DERP2.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 90256301.
RA CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
RA THOMAS W.R.;
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
RT plaque immunoassay."
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
RN [2]
RP PARTIAL SEQUENCE OF 18-57.
RX MEDLINE: 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATTS-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE: 98409423.
RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
RT "Tertiary structure of the major house dust mite allergen Der p 2:
RT sequential and structural homologies.";
RL Biochemistry 37:12707-12714(1998).
CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
DR PDB: 1A9V; 14-OCT-98
KW Allergen; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER P 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;
Query Match 100.0%; Score 106; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 1.34e+14;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 HEIKKVLVPGCHGS 41
 QY 1 HEIKKVLVPGCHGS 14
 |||||

RESULT 2
 ID DEF2_DERFA STANDARD; PRT; 146 AA.
 AC Q00855; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERF2.
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphidae;
 OC Dermatophagoides.
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 91291341.
 RA YUUKI T., OKUMURA Y., ANDO T., YAMAKAWA H., SUKO M., HAIDA M.,
 RA OKUDAIRA H.;
 RT "Cloning and expression of cDNA coding for the major house dust mite
 RT allergen Der f II in *Escherichia coli*.";
 RL Agric. Biol. Chem. 55:1233-1238(1991).
 RN [1]
 RN SEQUENCE OF 4-146 FROM N.A.
 RX MEDLINE; 94256850.
 RA OKUHARA H.;
 RT "Molecular biology of mite antigens.";
 RL Alerugi 43:435-440(1994).
 RN [3]
 RN DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE; 93283958.
 RA NISHIYAMA C., YUUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
 RT "determination of three disulfide bonds in a major house dust mite
 RT allergen, Der f II.";
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).
 RN [4]
 RN PARTIAL SEQUENCE OF 18-52.
 RX MEDLINE; 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATT'S-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [5]
 RN STRUCTURE BY NMR.
 RX MEDLINE; 98079068.
 RA ICHIKAWA S., HATANAKA H., YUUKI T., IWAMOTO N., KOJIMA S.,
 RA NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
 RT "Solution structure of Der f 2, the major mite allergen for atopic
 RT diseases.";
 RL J. Biol. Chem. 273:356-360(1998).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D10447; BAA01239.1;
 CC EMBL; D10448; BAA01240.1;
 CC EMBL; D10449; BAA01241.1;
 CC EMBL; S70378; AAB30829.1;

DR PIR; PS0417; PS0417,
 DR PDB; 1AHK; 08-APR-98.
 DR PDB; 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 FT VARIANT 93 93
 FT VARIANT 105 105 M -> V (IN CLONE 1).
 FT VARIANT 128 128 I -> A (IN CLONE 11).
 FT VARIANT 142 142 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMW (IN REF. 2).
 SQ SEQUENCE 146 AA; 72623F23 CRC32;
 Query Match 80.2%; Score 85; DB 1; Length 146;
 Best Local Similarity 78.6%; Pred. No. 2.19e-08;
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 28 NEIKKVLVPGCHGS 41
 QY 1 HEIKKVLVPGCHGS 14
 :|||||:|||||

RESULT 3
 ID YDID_ECOLI STANDARD; PRT; 548 AA.
 AC P38135; P76202; P76902; P76903;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL 60.8 KD PROTEIN IN AROD-PPSA INTERGENIC REGION.
 GN YDID.
 OS Escherichia coli
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 97426617.
 RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA AIBA H., BABA T., FUJITA K., HAYASHI K., HONJO A., HORIUCHI T.,
 RA IKEMOTO K., INADA T., ISONO K., ITOH T., KANAI K.,
 RA KASAI H., KASHIMOTO K., KIM S., KIMURA S., KITAGAWA M.,
 RA KITAKAWA M., MAKINO K., MASUDA S., MIKI T., MIZOBUCHI K., MORI H.,
 RA MOTOMURA K., NAKAMURA Y., NASHIMOTO H., NISHIO Y., OSHIMA T.,
 RA SAITO N., SAMPEI G., SEKI Y., TAGAMI H., TAKEMOTO K., WADA C.,
 RA YAMAMOTO Y., YANO M.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE OF 356-548 FROM N.A.
 RA HOLZSCHU D.L., MCELVER J.A., LIAO C.C., BERRY A.;
 RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN IDENTIFICATION.
 RX MEDLINE; 95075659.
 RA BORODOVSKY M., RUDD K.E., KOONIN E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome.";
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -1- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
 CC BINDING OF AMP TO THEIR SUBSTRATE. STRONG, TO 2,3-
 CC DIHYDROXYBENZOATE-AMP LIGASE.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB002065; AAC74771.1; ALT_INIT.
 CC EMBL; D90812; CAB21391.1; ALT_INIT.
 CC EMBL; D90813; CAB21407.1; ALT_INIT.
 CC EMBL; M69116; ; NOT_ANNOTATED_CDS.
 CC HSP; P08659; ILCI.
 CC ECGENE; EGI2357; YDID.
 CC PROSITE; PS00455; AMP_BINDING; 1.
 CC PFAM; PF00501; AMP-binding; 1.
 CC KW Hypothetical protein.
 CC SEQUENCE 548 AA; 60773 MW; CD545939 CRC32;
 CC -----
 CC Query Match 51.9%; Score 55; DB 1; Length 548;
 CC Best Local Similarity 50.0%; Pred. No. 1.13e+00;
 CC Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 CC -----
 CC Db 372 DARKTLPPGCG 383
 CC QY 2 EIRKLVPGCHG 13
 CC -----
 CC RESULT 4
 CC ID NIA2_HORVU STANDARD; PRT; 912 AA.
 CC AC P27969;
 CC DT 01-AUG-1992 (Rel. 23, Created)
 CC DT 01-AUG-1992 (Rel. 23, Last sequence update)
 CC DT 01-JUN-1994 (Rel. 29, Last annotation update)
 CC DE NITRATE REDUCTASE (EC 1.6.6.1) (NR) (FRAGMENT).
 CC OS Hordeum vulgare (Barley).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC OC Poaceae; Hordeum.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CV. STEPTOR; TISSUE=LEAF;
 CC RX MEDLINE; 91326031.
 CC RA SCHNORR K.M., JURICEK M., HUANG C., CULLEY D., KLEINHOF A.;
 CC RT "Analysis of barley nitrate reductase cDNA and genomic clones.";
 CC RL Mol. Gen. Genet. 227:411-416(1991).
 CC CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
 CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
 CC -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
 CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
 CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
 CC GROUP IS CALLED CYTOCHROME B-557.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
 CC N-TERMINAL DOMAIN.
 CC -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
 CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
 CC C-TERMINAL DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X57844; CAA40975.1; -
 CC PIR; S17454; RDBHNS.
 CC HSP; P17571; 2CND.
 CC PROSITE; PS00191; CYTOCHROME_B5; 1.
 CC PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
 CC KW Nitrate assimilation; Multigene family.

DR PFAM; PF00173; heme_1; 1.
 DR PFAM; PF00174; oxidored_molyb; 1.
 DR PFAM; PF00175; oxidored_fad; 1.
 DR PFAM; PF00970; Cyt_reductase; 1.
 KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
 KW Nitrate assimilation; Multigene family.
 FT NON_TER 1
 FT METAL 186 186 MOLYBDENUM-PTERIN (POTENTIAL).
 FT METAL 240 240 MOLYBDENUM-PTERIN (POTENTIAL).
 FT FT DISULFID 425 425 INTERCHAIN (POTENTIAL).
 FT BINDING 570 570 HEME LIGAND (BY SIMILARITY).
 FT BINDING 593 593 HEME LIGAND (BY SIMILARITY).
 SQ SEQUENCE 912 AA; 101467 MW; OB3FE625 CRC32;
 CC -----
 CC Query Match 51.9%; Score 55; DB 1; Length 912;
 CC Best Local Similarity 60.0%; Pred. No. 1.13e+00;
 CC Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC -----
 CC Db 291 RVLPGCIGG 300
 CC QY 5 KVLVPGCHGS 14
 CC -----
 CC RESULT 5
 CC ID NIA1_HORVU STANDARD; PRT; 915 AA.
 CC AC P27967;
 CC DT 01-AUG-1992 (Rel. 23, Created)
 CC DT 01-AUG-1992 (Rel. 23, Last sequence update)
 CC DT 01-JUN-1994 (Rel. 29, Last annotation update)
 CC DE NITRATE REDUCTASE (EC 1.6.6.1) (NR).
 CC OS Hordeum vulgare (Barley).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 CC OC Poaceae; Hordeum.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=CV. HIMALAYA; TISSUE=LEAF;
 CC RX MEDLINE; 91326031.
 CC RA SCHNORR K.M., JURICEK M., HUANG C., CULLEY D., KLEINHOF A.;
 CC RT "Analysis of barley nitrate reductase cDNA and genomic clones.";
 CC RL Mol. Gen. Genet. 227:411-416(1991).
 CC CC -!- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
 CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
 CC -!- CATALYTIC ACTIVITY: NADH + NITRATE = NAD(+) + NITRITE + H(2)O.
 CC -!- COFACTOR: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
 CC HEME IRON, AND MOLYBDENUM-PTERIN AS PROSTHETIC GROUPS. THE HEME
 CC GROUP IS CALLED CYTOCHROME B-557.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE
 CC N-TERMINAL DOMAIN.
 CC -!- SIMILARITY: CONTAINS A CYTOCHROME B5 FAMILY, HEME-BINDING DOMAIN.
 CC -!- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
 CC C-TERMINAL DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X57845; CAA40976.1; -
 CC PIR; S17453; RDBHNS.
 CC HSP; P17571; 2CND.
 CC PROSITE; PS00191; CYTOCHROME_B5; 1.
 CC PROSITE; PS00559; MOLYBDOPTERIN_EUK; 1.
 CC PFAM; PF00173; heme_1; 1.
 CC PFAM; PF00174; oxidored_molyb; 1.
 CC PFAM; PF00175; oxidored_fad; 1.
 CC PFAM; PF00970; Cyt_reductase; 1.
 KW Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
 KW Nitrate assimilation; Multigene family.

```

FT METAL 189 189 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 243 243 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 428 428 INTERCHAIN (POTENTIAL).
FT BINDING 573 573 HEME LIGAND (BY SIMILARITY).
FT BINDING 596 596 HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 915 AA; 101770 MW; F9319B2D CRC32;

Query Match 51.9%; Score 55; DB 1; Length 915;
Best Local Similarity 60.0%; Pred. No. 1.13e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 294 RVLPGCIGG 303
QY 5 KVLVPGCHS 14

RESULT 6
ID YJE3_YEAST STANDARD; PRT; 257 AA.
AC P47053;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOPHETICAL 29.2 KD PROTEIN IN GYP6-MHP1 INTERGENIC REGION.
GN YJL043W OR J1204.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA FOHL T.M., ALJINOVIC G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YK015C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49318; CAA89334.1; -
CC Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 257 AA; 29238 MW; A53A3882 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 257;
Best Local Similarity 72.7%; Pred. No. 1.89e+00;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 39 EILKSLVPKCH 49
QY 2 EIKKLVPGCH 12

RESULT 7
ID AMOA_AERYH STANDARD; PRT; 396 AA.
AC P23300;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PUTATIVE ISOCHORISMATE SYNTHASE (EC 5.4.99.6) (AMONABACTIN).
GN AMOA.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonas group;
OC Aeromonas.
RN [1]
RP SEQUENCE FROM N.A.
RA BARGHOOUTH S., PAYNE S.M., ARCENEAUX J.E., BYERS B.R.;
RX MEDLINE; 91317731.
RA "Cloning, mutagenesis, and nucleotide sequence of a siderophore
RT biosynthetic gene (amoA) from Aeromonas hydrophila.";
RL J. Bacteriol. 173:5121-5128(1991).

```

```

CC -1- CATALYTIC ACTIVITY: CHORISMATE - ISOCHORISMATE.
CC -1- PATHWAY: AMONABACTIN BIOSYNTHESIS. AMONABACTIN IS AN IRON-
CC CHELATING COMPOUND INVOLVED IN TRANSPORTING IRON FROM THE
CC BACTERIAL ENVIRONMENT INTO THE CELL CYTOPLASM.
CC -1- SIMILARITY: STRONG, TO OTHER ISOCHORISMATE SYNTHASES; WEAK, TO
CC TRPE AND PABB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63339; AAA21935.1; -
CC PIR; A40365; A40365
CC PFAM; PF00425; chorismate_bind; 1.
KW Isomerase; Iron transport.
SQ SEQUENCE 396 AA; 42074 MW; 7E8C32AD CRC32;

Query Match 50.9%; Score 54; DB 1; Length 396;
Best Local Similarity 54.5%; Pred. No. 1.89e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 247 EIRRVLPKCH 257
QY 2 EIKKLVPGCH 12

RESULT 8
ID YBP2_YEAST STANDARD; PRT; 397 AA.
AC P38226;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOPHETICAL 45.5 KD PROTEIN IN FIG1-GIP1 INTERGENIC REGION.
GN YBR042C OR YBR0412.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C;
RA ANDRE B., CZIEPLUCH C., HEIN C., JAUNIAUX J.-C., URRESTARAZU A.,
RA VISSERS S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE 1-ACYL-SN-GLYCEROL-3-PHOSPHATE
CC ACYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z35911; CAA84984.1; -
CC PIR; S45900; S45900.
CC PFAM; PF01553; Acyltransferase; 1.
KW Hypothetical protein; Phospholipid biosynthesis; Transferase;
KW Acyltransferase; Transmembrane.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 58 78 POTENTIAL.
FT TRANSMEM 118 138 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
SQ SEQUENCE 397 AA; 45515 MW; CCAD3795 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 397;
Best Local Similarity 60.0%; Pred. No. 1.89e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

```
Db 8 HKYKVVVPG 17
QY 1 HEKKVLVPG 10

RESULT 9
ID OBP_HSV7J STANDARD; PRT; 787 AA.
AC P52379;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE REPLICATION ORIGIN BINDING PROTEIN (OBP).
GN U73
OS Herpes simplex virus (type 7 / strain JI) (HSV7).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
RN [1]
RP SEQUENCE FROM N.A.
RA NICHOLAS J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLY INVOLVED IN DNA REPLICATION. BINDS THE ORIGIN
CC OF REPLICATION (ORI) (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL9,
CC HSV-1 53, AND VZV 51.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U43400; AAC54734.1; -
KW DNA replication; DNA-binding; ATP-binding.
FT NP_BIND 52 59
SQ SEQUENCE 787 AA; 90925 MW; 594C2D56 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 787;
Best Local Similarity 60.0%; Pred. No. 1.89e+00;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 447 FTKMLVQGGH 456
QY 3 IKKVLVPGCH 12

RESULT 10
ID DP3A_MYCTU STANDARD; PRT; 1184 AA.
AC Q10779;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE PROBABLE DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).
GN DNAE OR DNAEL OR RV1547 OR MTCY48.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROCHER K., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. II, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWEILL T., GENTLES S., HAMLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,
RA RUTTER S., SEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
```

```
RL Nature 393:537-544(1998).
CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N)
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74020; CAA98315.1; -
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 1184 AA; 129322 MW; D7EFB285 CRC32;

Query Match 50.9%; Score 54; DB 1; Length 1184;
Best Local Similarity 53.8%; Pred. No. 1.89e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 1172 DLKELGPGCLGS 1184
QY 2 EIKKVLVPGCHS 14

RESULT 11
ID NIFH_METVO STANDARD; PRT; 278 AA.
AC P06119;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE NITROGENASE IRON PROTEIN (EC 1.18.6.1) (NITROGENASE COMPONENT II)
DE (NITROGENASE REDUCTASE).
GN NIFH.
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RA SOUILLARD N., SIBOLD L.;
RT "Primary structure and expression of a gene homologous to nifH
RT (nitrogenase Fe protein) from the archaeobacterium Methanococcus
RT voltae.";
RL Mol. Gen. Genet. 203:21-28(1986).
CC -!- FUNCTION: THE KEY ENZYMATIC REACTIONS IN NITROGEN FIXATION ARE
CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.
CC -!- CATALYTIC ACTIVITY: 3 REDUCED FERREDOXIN + 6 H(+) + N(2) + N ATP -
CC 3 OXIDIZED FERREDOXIN + 2 NH(3) + N ADP + N ORTHOPHOSPHATE.
CC -!- COFACTOR: BINDS ONE 4Fe-4S CLUSTER PER DIMER.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE NIFH/FRXC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03777; CAA27407.1; -
PIR; S07313; NIMXVO.
HSP; P00459; INIP.
PROSITE; PS00692; NIFH_FRXC_2; 1.
PROSITE; PS00746; NIFH_FRXC_1; 1.
PFAM; PF00142; fer4_NifH; 1.
KW Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.
```


FT NP_BIND 8 15 ATP (POTENTIAL).
 FT METAL 94 94 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 130 130 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 278 AA; 30396 MW; F9866631 CRC32;

Query Match 50.0%; Score 53; DB 1; Length 278;

Best Local Similarity 87.5%; Pred. No. 3.12e+00;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 30 KKVLVGCG 37
 ||||| ||
 QY 4 KKVLVGCG 11

RESULT 12
 ID NIFH_METJA STANDARD; PRT; 279 AA.

AC Q58289;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE NITROGENASE IRON PROTEIN (EC 1.18.6.1) (NITROGENASE COMPONENT II)
 DE (NITROGENASE REDUCTASE).
 GN NIFH OR MJ0879.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

CC Methanococcus.

CC [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE; 96337999.

RA BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,
 RA SUTTON G.G., BLAKE J.A., FITZGERALD L.M., CLAYTON R.A., GOCAYNE J.D.,
 RA KERLAVAGE A.R., DOUGHERTY B.A., TOMB J.-F., ADAMS M.D., REICH C.I.,
 RA OVERBECK R., KIRKNESS E.F., WEINSTOCK K.G., MERRICK J.M., GLODEK A.,
 RA SCOTT J.L., GEORGHAN N.S.M., WEIDMAN J.E., FUHRMANN J.L., NGUYEN D.,
 RA UTTERBACK T.R., KELLEY J.M., PETERSON J.D., SADOW P.W., HANNA M.C.,
 RA COTTON M.D., ROBERTS K.M., HURST M.A., KAINE B.P., BORODOVSKY M.,
 RA KLENK H.-P., FRASER C.M., SMITH H.O., WOESE C.R., VENTER J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii";
 RL Science 273:1058-1073 (1996).

CC -1- FUNCTION: THE KEY ENZYMATIC REACTIONS IN NITROGEN FIXATION ARE
 CC CATALYZED BY THE NITROGENASE COMPLEX, WHICH HAS 2 COMPONENTS: THE
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 3 REDUCED FERREDOXIN + 6 H(+) + N(2) + N ATP =
 CC 3 OXIDIZED FERREDOXIN + 2 NH(3) + N ADP + N ORTHOPHOSPHATE.
 CC -1- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NIFH/FRXC FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; U67532; AAB98883.1; -;
 CC HSP; P04559; IN2C.
 CC TIGR; MJ0879; -;
 CC DR PROSITE; PS00692; NIFH_FRXC_2; 1.
 CC DR PROSITE; PS00746; NIFH_FRXC_1; 1.
 CC DR PFAM; PF00142; fer4_NifH; 1.
 CC DR NP_BIND 8 15
 CC KX Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.
 CC FT NP_BIND 8 15
 CC FT METAL 94 94 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 CC FT METAL 130 130 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 279 AA; 30167 MW; 00B2854B CRC32;

Query Match 50.0%; Score 53; DB 1; Length 279;

Best Local Similarity 87.5%; Pred. No. 3.12e+00;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 30 KKVLVGCG 37
 ||||| ||
 QY 4 KKVLVGCG 11

RESULT 13

ID OAT_YEAST STANDARD; PRT; 424 AA.

AC P07991;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE ORNITHINE AMINOTRANSFERASE (EC 2.6.1.13) (ORNITHINE--OXO-ACID

DE AMINOTRANSFERASE).

GN CAR2 OR CARGO OR YLR438W OR L9753.2.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

CC Saccharomycetaceae; Saccharomyces.

CC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 88055042.

RA DEGOLS G.;

RT "Functional analysis of the regulatory region adjacent to the carb

RT gene of Saccharomyces cerevisiae. Nucleotide sequence, gene fusion

RT experiments and cis-dominant regulatory mutation analysis.";

RL Eur. J. Biochem. 169:193-200 (1987).

CC [2]

RP SEQUENCE FROM N.A.

RC STRAIN-S288C / AB972;

RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU 2.,

RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,

RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,

RA JOHNSTON D., JOHNSTON L., LANGSTON Y., LATREILLE P., LE T.,

RA MARDIS E., MENEZES S., MILLER N., NHAN M., PAULEY A., PELUSO D.,

RA RIEKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,

RA WILCOX L., WOLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

CC [3]

RP SEQUENCE OF 1-55 FROM N.A.

RC STRAIN-SIGMA 1278B;

RX MEDLINE; 87246605.

RA DEGOLS G., JAUNIAUX J.-C., WIAME J.M.;

RT "Molecular characterization of transposable-element-associated

RT mutations that lead to constitutive L-ornithine aminotransferase

RT expression in Saccharomyces cerevisiae.";

RL Eur. J. Biochem. 165:389-396 (1987).

CC -1- CATALYTIC ACTIVITY: L-ORNITHINE + A 2-OXO ACID - L-GLUTAMATE

CC 5-SEMIALDEHYDE + AN L-AMINO ACID.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT

CC AMINOTRANSFERASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; X06790; CAA29947.1; -;

CC EMBL; U21094; AAB67514.1; -;

CC EMBL; X05571; CAA29081.1; -;

CC PIR; S00181; XNBYO.

CC HSP; P04181; 1OAT.

CC SGD; L0000217; CAR2.

CC PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.

CC PFAM; PF00202; aminotran_3; 1.

CC Transferase; Aminotransferase; Pyridoxal phosphate.

KW TRANSFERASE; AMINOTRANSFERASE; PYRIDOXAL PHOSPHATE (BY SIMILARITY).

FT BINDING 272 272 PYRIDOXAL PHOSPHATE (IN REF. 1 AND 3).

FT CONFLICT 8 8 MISSING (IN REF. 1 AND 3).

FT CONFLICT 38 38 V -> L (IN REF. 1 AND 3).

```

FT CONFLICT 99 99 F -> S (IN REF. 1).
FT CONFLICT 212 212 G -> R (IN REF. 1).
FT CONFLICT 385 385 H -> Q (IN REF. 1).
SQ SEQUENCE 424 AA; 46086 MW; 3F432867 CRC32;

Query Match
Best Local Similarity 50.0%; Score 53; DB 1; Length 424;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 288 HDIMSCFTFGSHGS 301
QY 1 HEIKVLVPCGHS 14

RESULT 14
ID ETFD_YEAST STANDARD; PRT; 631 AA.
AC Q08822;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-UBIQUINONE OXIDOREDUCTASE
DE PRECURSOR (EC 1.5.1.1) (ETF-QO) (ETF-UBIQUINONE OXIDOREDUCTASE) (ETF
DE DEHYDROGENASE) (ELECTRON-TRANSFERRING-FLAVOPROTEIN DEHYDROGENASE).
GN YOR356W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
[1]
RN SEQUENCE FROM N.A.
RA DELIUS H., HEILING U., HOFMANN B.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACCEPTS ELECTRONS FROM ETF AND REDUCES UBIQUINONE (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: REDUCED ETF + UBIQUINONE -> ETF + UBIQUINOL.
CC -!- COFACTOR: FAD AND A 4FE-4S CLUSTER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ETF-QO / FIXC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z75264; CAA99685.1;
KW Oxidoreductase; Electron transport; Flavoproteins; FAD; Iron-sulfur;
KW 4Fe-4S; Mitochondrion; Transist peptide; Ubiquinone.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 631 PROBABLE ELECTRON TRANSFER FLAVOPROTEIN-
FT UBIQUINONE OXIDOREDUCTASE.
FT NP_BIND 65 79 FAD (ADP PART) (POTENTIAL).
FT METAL 574 574 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 600 600 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 603 603 IRON-SULFUR (4FE-4S) (POTENTIAL).
FT METAL 606 606 IRON-SULFUR (4FE-4S) (POTENTIAL).
SQ SEQUENCE 631 AA; 69634 MW; 6CE9A3B3 CRC32;

Query Match
Best Local Similarity 50.0%; Score 53; DB 1; Length 631;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 242 VLAEGCHGS 250
QY 6 VLVPGCHGS 14

RESULT 15
ID LCAT_PIG STANDARD; PRT; 188 AA.
AC P30930;
DT 01-JUL-1993 (Rel. 26, Created)

```

```

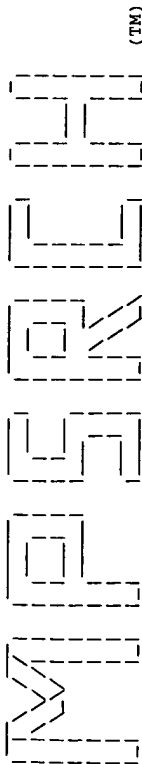
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-
DE CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
DE ACYLTRANSFERASE) (FRAGMENTS).
GN LCAT.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE.
RC TISSUE-PLASMA;
RX MEDLINE; 90075705.
RA YUEKSEL K.U., PARK Y.B., JUNG J., GRACY R.W., LACKO A.G.;
RT "Studies on the structure of lecithin:cholesterol acyltransferase
RT (LACT) -- comparisons of the active site region and secondary
RT structure of the human and the porcine enzymes.";
RL Comp. Biochem. Physiol. 94B:389-394(1989).
RN [2]
RP SEQUENCE OF 97-106 AND 142-154.
RC TISSUE-PLASMA;
RX MEDLINE; 87156771.
RA PARK Y.B., YUEKSEL K.U., GRACY R.W., LACKO A.G.;
RT "The catalytic center of lecithin:cholesterol acyltransferase:
RT isolation and sequence of diisopropyl fluorophosphate-labeled
RT peptides.";
RL Biochem. Biophys. Res. Commun. 143:360-363(1987).
CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL -> STEROL ESTER +
CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
CC ACT AS ACCEPTOR).
CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
CC THIS ENZYME.
CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
DR PIR; PLO153.
DR PIR; A29544; A29544.
DR PIR; B29544; B29544.
DR PROSITE; PS00120; LIPASE_SER; PARTIAL.
KW Transferase; Acyltransferase; Lipid metabolism; Glycoprotein.
FT CARBOHYD 20 20
FT NON_CONS 34 35
FT NON_CONS 44 45
FT NON_CONS 60 61
FT NON_CONS 66 67
FT NON_CONS 77 78
FT NON_CONS 84 85
FT NON_CONS 96 97
FT NON_CONS 106 107
FT NON_CONS 115 116
FT NON_CONS 154 155
FT NON_TER 188 188
SQ SEQUENCE 188 AA; 21232 MW; 876EB48E CRC32;

Query Match
Best Local Similarity 49.1%; Score 52; DB 1; Length 188;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 26 ILVPGCLGN 34
QY 6 VLVPGCHGS 14

```

Search completed: Sun Apr 2 17:52:43 2000
Job time : 32 secs.



 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:51:06 2000; MasPar time 44.19 Seconds
 Tabular output not generated. 14.944 Million cell updates/sec

Title: >US-09-362-731-2
 Description: (1-14) from US09362731.ppe
 Perfect Score: 106
 Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: PAM 150
 Gap 15

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: pir62
 1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 25.741; Variance 32.196; scale 0.800

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	100%	146	2 A60381	allergen Der p II pre	2.37e-12
2	85	80.2	129	2 A61501	allergen Der f II - h	6.62e-07
3	85	80.2	129	2 JU0394	allergen Der f II (pf	6.62e-07
4	85	80.2	138	2 A61241	allergen Der f II pre	6.62e-07
5	85	80.2	138	2 A61241	allergen Der f II pre	6.62e-07
6	57	53.8	77	2 A69906	hypothetical protein	1.71e+00
7	55	51.9	250	2 F72247	methionine aminopepti	4.32e+00
8	55	51.9	566	2 E64928	probable acid--CoA li	4.32e+00
9	55	51.9	912	1 RBHNS	nitrate reductase (NA	4.32e+00
10	55	51.9	915	1 RBHNS	nitrate reductase (NA	4.32e+00
11	54	50.9	237	2 S56815	hypothetical protein	6.80e+00
12	54	50.9	396	1 A40365	siderophore biosynthe	6.80e+00
13	54	50.9	397	2 S45900	probable membrane pro	6.80e+00
14	54	50.9	482	1 JE09054	DNA-directed DNA poly	6.80e+00
15	54	50.9	960	2 JE0356	gamma-aminobutyric ac	6.80e+00
16	54	50.9	1184	2 H70761	probable dnapolymeras	6.80e+00
17	53	50.0	278	1 NIMXVO	nitrogenase (EC 1.18.	1.06e+01
18	53	50.0	279	2 G64409	nitrogenase (EC 1.18.	1.06e+01
19	53	50.0	424	1 XNBYO	ornithine--oxo-acid t	1.06e+01
20	53	50.0	426	2 T04527	hypothetical protein	1.06e+01
21	53	50.0	442	2 S50062	cell wall glycoprotei	1.06e+01
22	53	50.0	631	2 S67268	hypothetical protein	1.06e+01
23	53	50.0	632	2 S64786	hypothetical protein	1.06e+01

24 52 49.1 110 2 B72730 hypothetical protein 1.65e+01
 25 52 49.1 141 2 B30868 hypothetical protein 1.65e+01
 26 52 49.1 179 2 C60328 conserved hypothetical 1.65e+01
 27 52 49.1 182 2 E72312 phosphatidylcholine-- 1.65e+01
 28 52 49.1 188 2 PL0153 thymidylate synthase 1.65e+01
 29 52 49.1 328 2 S73842 phosphatidylcholine-- 1.65e+01
 30 52 49.1 438 1 XXMSN phosphatidylcholine-- 1.65e+01
 31 52 49.1 440 2 JCL502 phosphatidylcholine-- 1.65e+01
 32 52 49.1 440 1 XXHUN phosphatidylcholine-- 1.65e+01
 33 52 49.1 1194 2 S70415 DNA-directed RNA poly 1.65e+01
 34 51 48.1 48 2 S55783 laminin gamma chain - 2.55e+01
 35 51 48.1 141 2 C69018 methyl viologen-reduc 2.55e+01
 36 51 48.1 248 2 F72262 ubiquinone/menaquinon 2.55e+01
 37 51 48.1 249 2 E71649 hypothetical protein 2.55e+01
 38 51 48.1 286 2 S62370 serine-type carboxype 2.55e+01
 39 51 48.1 415 2 T07706 hypothetical protein 2.55e+01
 40 51 48.1 440 1 XXRTN phosphatidylcholine-- 2.55e+01
 41 51 48.1 505 2 T15159 hypothetical protein 2.55e+01
 42 51 48.1 654 2 T10521 beta-glucosidase (EC 2.55e+01
 43 51 48.1 655 2 C65243 hypothetical 70.1 kD 2.55e+01
 44 51 48.1 1607 1 MMWSB2 laminin gamma-1 chain 2.55e+01
 45 51 48.1 2499 1 A30788 mannose 6-phosphate r 2.55e+01

ALIGNMENTS

RESULT 1

ENTRY A60381 #type complete
 TITLE allergen Der p II precursor - house-dust mite
 (Dermatophagoides pteronyssinus)
 ORGANISM #formal_name Dermatophagoides pteronyssinus
 DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 13-Sep-1998
 A60381
 REFERENCE Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;
 Stewart, G.A.; Thomas, W.R.
 #journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
 #title Isolation of cDNA coding for the major mite allergen Der p II
 by IGE plaque immunoassay.
 #cross-references MUID:90256301
 #accession A60381
 #status not compared with conceptual translation
 #molecule_type mRNA
 #residues 1-146 #label CHU
 CLASSIFICATION #superfamily allergen Der p II
 FEATURE
 1-17 #domain signal sequence #status predicted #label SIG\
 18-146 #product allergen Der p II #status predicted #label MAT
 SUMMARY #length 146 #molecular-weight 15999 #checksum 25

Query Match 100.0%; Score 106; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 2.37e-12;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 28 HEIKKVLVPGCHGS 41

QY 1 HEIKKVLVPGCHGS 14

RESULT 2

ENTRY A61501 #type fragment
 TITLE allergen Der f II - house-dust mite (Dermatophagoides
 farinae) (fragment)
 ORGANISM #formal_name Dermatophagoides farinae
 DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
 A61501

REFERENCE A61501
 #journal Trudinger, M.; Chua, K.Y.; Thomas, W.R.
 Clin. Exp. Allergy (1991) 21:33-37

#title cDNA encoding the major mite allergen Der f II.
 #cross-references MUID:91215495

```

#accession A61501
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-129 #label TRU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476

Query Match 80.2% Score 85; DB 2; Length 129;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 11 NEIKKVMVDCGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
:|||||:|||||

RESULT 3
ENTRY #type complete
TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
ACCESSIONS JU0394
REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Okudaira, H.
#journal Agric. Biol. Chem. (1991) 55:1233-1238
#title Cloning and expression of cDNA coding for the major house dust mite allergen Der f II in Escherichia coli.
#cross-references MUID:91291341
#accession JU0394
#molecule_type mRNA
#residues 1-129 #label YUU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 80.2% Score 85; DB 2; Length 129;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 11 NEIKKVMVDCGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
:|||||:|||||

RESULT 4
ENTRY #type fragment
TITLE allergen Der f II precursor - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS A61241; PS0417
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession A61241
#molecule_type mRNA
#residues 1-138 #label YUU
#note part of this sequence, including the amino end of the mature protein, was confirmed by protein sequencing
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
10-138 #product allergen Der f II #status experimental #label MAT
SUMMARY #length 138 #checksum 2894

Query Match 80.2% Score 85; DB 2; Length 138;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 NEIKKVMVDCGCHGS 33
QY 1 HEIKKVLVPGCHGS 14
:|||||:|||||

RESULT 5
ENTRY #type fragment
TITLE allergen Der f II precursor - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change 13-Sep-1998
ACCESSIONS B61241; JU0395
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.; Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession B61241
#molecule_type mRNA
#residues 1-138 #label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
10-138 #product allergen Der f II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 80.2% Score 85; DB 2; Length 138;
Best Local Similarity 78.6% Pred. No. 6.62e-07;
Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 NEIKKVMVDCGCHGS 33
QY 1 HEIKKVLVPGCHGS 14
:|||||:|||||

RESULT 6
ENTRY #type complete
TITLE hypothetical protein yojC - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
ACCESSIONS A69906
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Gali, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kashahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;

```

```

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession A69906
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-77 #label KUN
#cross-references GB:299114; GB:AL009126; NID:g2634230; PID:el185421;
PID:g2634342
#experimental_source strain 168
GENETICS
#gene yojC
#length 77 #molecular-weight 8518 #checksum 9771
Query Match 53.8%; Score 57; DB 2; Length 77;
Best Local Similarity 66.7%; Pred. No. 1.71e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 14 RTVLVPACH 22
: |||||
QY 4 KRVLPVPGCH 12

RESULT 7
ENTRY #type complete
TITLE methionine aminopeptidase - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
23-Jul-1999
ACCESSION F72247
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession F72247
#status preliminary
#molecule_type DNA
#residues 1-250 #label ARN
#cross-references GB:AE001798; GB:AE000512; NID:g4982033; PID:g4982042;
TIGR:TM1478
#experimental_source strain MSB8
GENETICS
#gene TM1478
#length 250 #molecular-weight 27489 #checksum 7507
CLASSIFICATION #superfamily Escherichia coli methionyl aminopeptidase
SUMMARY
Query Match 51.9%; Score 55; DB 2; Length 250;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 25 REVRKIVPG 34

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession A69906
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-77 #label KUN
#cross-references GB:299114; GB:AL009126; NID:g2634230; PID:el185421;
PID:g2634342
#experimental_source strain 168
GENETICS
#gene yojC
#length 77 #molecular-weight 8518 #checksum 9771
Query Match 53.8%; Score 57; DB 2; Length 77;
Best Local Similarity 66.7%; Pred. No. 1.71e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 14 RTVLVPACH 22
: |||||
QY 4 KRVLPVPGCH 12

RESULT 7
ENTRY #type complete
TITLE methionine aminopeptidase - Thermotoga maritima (strain MSB8)
ORGANISM #formal_name Thermotoga maritima
DATE 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change
23-Jul-1999
ACCESSION F72247
REFERENCE A72200
#authors Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson,
R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Nelson,
W.C.; Ketchum, K.A.; McDonald, L.; Utterback, T.R.; Malek,
J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.; Cotton,
M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
Heidelberg, J.; Sutton, G.G.; Fleischmann, R.D.; White, O.;
Salzberg, S.L.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
#journal Nature (1999) 399:323-329
#title Evidence for lateral gene transfer between Archaea and
Bacteria from genome sequence of Thermotoga maritima.
#cross-references MUID:99287316
#accession F72247
#status preliminary
#molecule_type DNA
#residues 1-250 #label ARN
#cross-references GB:AE001798; GB:AE000512; NID:g4982033; PID:g4982042;
TIGR:TM1478
#experimental_source strain MSB8
GENETICS
#gene TM1478
#length 250 #molecular-weight 27489 #checksum 7507
CLASSIFICATION #superfamily Escherichia coli methionyl aminopeptidase
SUMMARY
Query Match 51.9%; Score 55; DB 2; Length 250;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 25 REVRKIVPG 34

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Taccioni, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium
Bacillus subtilis.
#cross-references MUID:98044033
#accession A69906
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-566 #label BLAT
#cross-references GB:AE000265; GB:U00096; NID:g2367122; PID:gl787993;
UWGP:b1701
#experimental_source strain K-12, substrain MG1655
GENETICS
#gene yidJ
#length 566 #molecular-weight 62759 #checksum 7145
Query Match 51.9%; Score 55; DB 2; Length 566;
Best Local Similarity 50.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 390 DAKTLPPGCEG 401
: |||||
QY 2 EIKKVLVPVPGCH 13

RESULT 9
ENTRY #type fragment
TITLE nitrate reductase (NADH) (EC 1.6.6.1) - barley (cv. Steptoe)
(fragment)
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
11-Jun-1999
ACCESSION S17454
REFERENCE S17453
#authors Schnorr, K.M.; Juricek, M.; Huang, C.; Culley, D.; Kleinhofs,
A.
#journal Mol. Gen. Genet. (1991) 227:411-416
#title Analysis of barley nitrate reductase cDNA and genomic clones.
#cross-references MUID:91326031
#accession S17454
#status nucleic acid sequence not shown; translation not shown
#molecule_type mRNA
#residues 1-912 #label SCH
#cross-references EMBL:X57844; NID:g19044; PID:CAA00975.1; PID:g19045
#note the nucleotide sequence was submitted to the EMBL Data
Library, February 1991
GENETICS
#map_position 6
CLASSIFICATION #superfamily nitrate reductase (NADH); cytochrome b5 core
homology; cytochrome-b5 reductase homology;
molybdopterin-binding domain homology

```

```

QY 1 HEIKVLVPG 10
: |||||

```

```

RESULT 8

```

```

ENTRY #type complete

```

```

TITLE probable acid-CoA ligase (EC 6.2.1.-) yidJ - Escherichia

```

```

ORGANISM coli

```

```

DATE #formal_name Escherichia coli

```

```

ACCESSIONS E64928

```

```

REFERENCE E64928

```

```

#authors A64720

```

```

Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;

```

```

Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;

```

```

Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;

```

```

Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,

```

```

Y.

```

```

#journal Science (1997) 277:1453-1462

```

```

#title The complete genome sequence of Escherichia coli K-12.

```

```

#cross-references MUID:97426617

```

```

#accession E64928

```

```

#status preliminary; nucleic acid sequence not shown;

```

```

translation not shown

```

```

#molecule_type DNA

```

```

#residues 1-566 #label BLAT

```

```

#cross-references GB:AE000265; GB:U00096; NID:g2367122; PID:gl787993;

```

```

UWGP:b1701

```

```

#experimental_source strain K-12, substrain MG1655

```

```

GENETICS

```

```

#gene yidJ

```

```

CLASSIFICATION #superfamily 4-coumarate--CoA ligase; acetate--CoA ligase

```

```

homology

```

```

KEYWORDS acid-thiol ligase; ATP

```

```

FEATURE

```

```

89-349

```

```

SUMMARY

```

```

Query Match 51.9%; Score 55; DB 2; Length 566;

```

```

Best Local Similarity 50.0%; Pred. No. 4.32e+00;

```

```

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

Db 390 DAKTLPPGCEG 401

```

```

: |||||

```

```

QY 2 EIKKVLVPVPGCH 13

```

```

RESULT 9

```

```

ENTRY #type fragment

```

```

TITLE nitrate reductase (NADH) (EC 1.6.6.1) - barley (cv. Steptoe)

```

```

(fragment)

```

```

ORGANISM #formal_name Hordeum vulgare #common_name barley

```

```

DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change

```

```

11-Jun-1999

```

```

ACCESSIONS S17454

```

```

REFERENCE S17453

```

```

#authors Schnorr, K.M.; Juricek, M.; Huang, C.; Culley, D.; Kleinhofs,

```

```

A.

```

```

#journal Mol. Gen. Genet. (1991) 227:411-416

```

```

#title Analysis of barley nitrate reductase cDNA and genomic clones.

```

```

#cross-references MUID:91326031

```

```

#accession S17454

```

```

#status nucleic acid sequence not shown; translation not shown

```

```

#molecule_type mRNA

```

```

#residues 1-912 #label SCH

```

```

#cross-references EMBL:X57844; NID:g19044; PID:CAA00975.1; PID:g19045

```

```

#note the nucleotide sequence was submitted to the EMBL Data

```

```

Library, February 1991

```

```

GENETICS

```

```

#map_position 6

```

```

CLASSIFICATION #superfamily nitrate reductase (NADH); cytochrome b5 core

```

```

homology; cytochrome-b5 reductase homology;

```

```

molybdopterin-binding domain homology

```

KEYWORDS chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron; metalloprotein; molybdenum; molybdopterin; NAD; nitrate assimilation; oxidoreductase; phosphoprotein

FEATURE 82-474 #domain molybdopterin-binding domain homology #label PCO\

535-609 #domain cytochrome b5 core homology #label CB5\
658-912 #domain cytochrome-b5 reductase homology #label CBR\
186 #binding_site molybdopterin (Cys) (covalent) #status predicted\
425 #disulfide_bonds interchain #status predicted\
570,593 #binding_site heme iron (His) (axial ligands) #status predicted\
722,884 #binding_site NAD (Lys, Cys) #status predicted\
763 #binding_site FAD (Tyr) #status predicted
SUMMARY #length 912 #checksum 1642

Query Match 51.9%; Score 55; DB 1; Length 912;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 291 RVLPGCIGG 300
QY 5 KVLVPGCHGS 14
:|||||:

RESULT 10 RDBNH #type complete
ENTRY nitrate reductase (NADH) (EC 1.6.6.1) - barley (cv. Himalaya)
TITLE #formal_name Hordeum vulgare #common_name barley
ORGANISM 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
DATE 11-Jun-1999

ACCESSIONS S17453
REFERENCE S17453
#authors Schnorr, K.M.; Juricek, M.; Huang, C.; Culley, D.; Kleinhofs, A.

#journal Mol. Gen. Genet. (1991) 227:411-416
#title Analysis of barley nitrate reductase cDNA and genomic clones.
#cross-references EMBL:X57845; NID:g18993; PIDN:CAA40976.1; PID:g18994
#accession #cross-references M17453

#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-915 #label SCH
#cross-references EMBL:X57845; NID:g18993; PIDN:CAA40976.1; PID:g18994
#note the nucleotide sequence was submitted to the EMBL Data Library, February 1991

GENETICS #map_position 6
#introns 391/1
CLASSIFICATION #superfamily nitrate reductase (NADH); cytochrome b5 core homology; cytochrome-b5 reductase homology;
molybdopterin-binding domain homology

KEYWORDS chromoprotein; electron transfer; FAD; flavoprotein; heme; homodimer; iron; metalloprotein; molybdenum; molybdopterin; NAD; nitrate assimilation; oxidoreductase; phosphoprotein

FEATURE 85-477 #domain molybdopterin-binding domain homology #label PCO\
538-612 #domain cytochrome b5 core homology #label CB5\
661-915 #domain cytochrome-b5 reductase homology #label CBR\
189 #binding_site molybdopterin (Cys) (covalent) #status predicted

428 #disulfide_bonds interchain #status predicted\
573,596 #binding_site heme iron (His) (axial ligands) #status predicted\
725,887 #binding_site NAD (Lys, Cys) #status predicted\
766 #binding_site FAD (Tyr) #status predicted
SUMMARY #length 915 #molecular-weight 101770 #checksum 3558

Query Match 51.9%; Score 55; DB 1; Length 915;
Best Local Similarity 60.0%; Pred. No. 4.32e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 294 RVLPGCIGG 303
QY 5 KVLVPGCHGS 14
:|||||:

RESULT 11

ENTRY S56815 #type complete
TITLE hypothetical protein YJL043w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein J1204
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 14-Nov-1997

ACCESSIONS S56815
REFERENCE S56793
#authors Pohl, T.M.; Aljinovic, G.
#submission submitted to the Protein Sequence Database, September 1995
#accession S56815
#molecule_type DNA
#residues 1-257 #label TOV
#cross-references EMBL:249318; NID:g1008170; MIPS:YJL043w

GENETICS #map_position 10L
SUMMARY #length 257 #molecular-weight 29238 #checksum 4555

Query Match 50.9%; Score 54; DB 2; Length 257;
Best Local Similarity 72.7%; Pred. No. 6.80e+00;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 39 EILKSLVPKCH 49
QY 2 EIKKVLVPKCH 12
:|||||:

RESULT 12

ENTRY A40365 #type complete
TITLE siderophore biosynthetic protein amoA - Aeromonas hydrophila
ORGANISM #formal_name Aeromonas hydrophila
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS A40365
REFERENCE A40365
#authors Barghouti, S.; Payne, S.M.; Arceneaux, J.E.L.; Byers, B.R.
#journal J. Bacteriol. (1991) 173:5121-5128
#title Cloning, mutagenesis, and nucleotide sequence of a siderophore biosynthetic gene (amoA) from Aeromonas hydrophila.

#cross-references M17731
#accession A40365
#status preliminary
#molecule_type DNA
#residues 1-396 #label BAR
#cross-references GB:M63339
#note the authors translated the codon GAG for residue 393 as Gly

CLASSIFICATION #superfamily isochorismate synthase
SUMMARY #length 396 #molecular-weight 42042 #checksum 6097

Query Match 50.9%; Score 54; DB 1; Length 396;
Best Local Similarity 54.5%; Pred. No. 6.80e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 247 EIRRVLTPCR 257
QY 2 EIKKVLVPKCH 12
:|||||:

RESULT 13

ENTRY S45900 #type complete
TITLE probable membrane protein YBR042c - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein YBR0412
ORGANISM #formal_name Saccharomyces cerevisiae

```

DATE          26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change
ACCESSIONS    S45900
REFERENCE      Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.;
#authors      Urrestazu, A.; Vissers, S.
#submission   submitted to the Protein Sequence Database, August 1994
#accession    S45900
#molecule_type DNA
##residues    1-397 ##label AND
##cross-references EMBL:Z35911; NID:G536265; PID:CAA94984.1;
#experimental_source strain S288C
GENETICS
#map_position 2R
CLASSIFICATION #superfamily probable membrane protein YBR042C
KEYWORDS       transmembrane protein
FEATURE
12-37          #domain transmembrane #status predicted #label TM1\
55-77          #domain transmembrane #status predicted #label TM2\
134-150        #domain transmembrane #status predicted #label TM3\
372-390        #domain transmembrane #status predicted #label TM4\
SUMMARY        #length 397 #molecular-weight 45515 #checksum 7874
Query Match.   50.9%; Score 54; DB 2; Length 397;
Best Local Similarity 60.0%; Pred. No. 6.80e+00;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 8 HKVRKVVP 17
QY 1 HEIKKVLVP 10
RESULT 14
ENTRY      B69054 #type complete
TITLE      DNA-directed DNA polymerase (EC 2.7.7.7) delta small chain -
ORGANISM   Methanobacterium thermoautotrophicum (strain Delta H)
DATE       05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS B69054
REFERENCE   A69000
#authors    Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
            Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
            Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
            Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
            Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
            A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
            McDougall, S.; Shimer, G.; Goyal, A.; Pietrokovski, S.;
            Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
            J.; Reeve, J.N.
#journal    J. Bacteriol. (1997) 179:7135-7155
#title      Complete genome sequence of Methanobacterium
            thermoautotrophicum Delta H: functional analysis and
            comparative genomics.
#cross-references MUID:98037514
#accession   B69054
#status      nucleic acid sequence not shown; translation not shown
#residues    1-482 ##label MTH
##cross-references GB:AE000903; GB:AE000666; NID:g2622514;
            PID:AA85882.1; PID:g2622517
#experimental_source strain Delta H
GENETICS
#gene       MTH1405
CLASSIFICATION #superfamily DNA-directed DNA polymerase delta small chain;
            phosphoesterase core homology
KEYWORDS     metalloprotein; nucleotidyltransferase
FEATURE
223-321      #domain phosphoesterase core homology #label PEC
SUMMARY      #length 482 #molecular-weight 54375 #checksum 3013
Query Match  50.9%; Score 54; DB 1; Length 482;

```

```

Best Local Similarity 53.8%; Pred. No. 6.80e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
Db 310 DIKIVMIPGRHDS 322
QY 2 EIRKVLVP 14
RESULT 15
ENTRY      JE0356 #type complete
TITLE      gamma-aminobutyric acid receptor B precursor - human
ALTERNATE_NAMES GABA(B) receptor
ORGANISM      #formal_name Homo sapiens #common_name man
DATE          05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change
ACCESSIONS   JE0356
REFERENCE     Grifa, A.; Totaro, A.; Rommens, J.M.; Carella, M.; Roetto,
            A.; Borgato, L.; Zelante, L.; Gasparini, P.;
            Biochem. Biophys. Res. Commun. (1998) 250:240-245
            GABA (gamma-amino-butyric acid) neurotransmission:
            Identification and fine mapping of the human GABAB receptor
            gene.
#cross-references MUID:98440782
#accession   JE0356
#molecule_type mRNA
##residues  1-960 ##label GRI
##cross-references GB:Y11044; NID:g2826760
            this ORF is not annotated in GenBank entry HSGTHLAL,
            release 109
GENETICS
#map_position 6p21.3-6p21.3
KEYWORDS     glycoprotein; neurotransmitter receptor; transmembrane
            protein
FEATURE
1-11         #domain signal sequence #status predicted #label SIG\
12-960       #product gamma-aminobutyric acid receptor B #status
            predicted #label MAR\
590-613      #domain transmembrane #status predicted #label TM1\
627-654      #domain transmembrane #status predicted #label TM2\
666-687      #domain transmembrane #status predicted #label TM3\
709-730      #domain transmembrane #status predicted #label TM4\
767-788      #domain transmembrane #status predicted #label TM5\
803-825      #domain transmembrane #status predicted #label TM6\
831-856      #domain transmembrane #status predicted #label TM7\
23,83,439,481,501,
513,630      #binding_site carbohydrate (Asn) (covalent) #status
            Predicted
SUMMARY      #length 960 #molecular-weight 108148 #checksum 3766
Query Match  50.9%; Score 54; DB 2; Length 960;
Best Local Similarity 66.7%; Pred. No. 6.80e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 237 IKIILMPGC 245
QY 3 IKKVLVPGC 11
Search completed: Sun Apr 2 17:51:55 2000
Job time : 49 secs.

```

(TM)

Result No.	Score	Match	Query Length	DB Length	ID	Description	Pred. No.
1	101	45.1	145	5	096430	GROUP 2 ALLERGEN EUR M	1.10e+06
2	74	33.0	635	14	009496	K5 LYASE.	2.64e-01
3	73	32.0	205	2	092CW7	SC02 PROTEIN PRECURSOR	4.01e-01
4	72	32.1	210	10	039969	FUCOXANTHIN CHLOROPHYL	6.07e-01
5	72	32.1	471	5	09Y1X6	STPRAA (EC 3.1.3.48) (6.07e-01
6	70	31.3	208	5	022814	T76C5.1 PROTEIN.	1.37e+00
7	70	31.3	393	1	039542	393AA LONG HYPOTHETICA	1.37e+00
8	70	31.3	841	10	092PP5	PUTATIVE POL POLYPROTE	1.37e+00
9	68	30.4	189	2	092BN6	PUTATIVE TRANSCRIPTION	3.07e+00
10	68	30.4	224	14	P89671	CAPSID PROTEIN (FRAGME	3.07e+00
11	67	29.9	291	2	092ND1	HEXOXYMETHYLBILANE SY	4.55e+00
12	67	28.9	486	3	Q05838	HEXOKINASE PI (HKX2) .	4.55e+00
13	67	29.9	517	10	Q41268	OSNRAMP1.	4.55e+00
14	67	28.9	692	2	092K24	ELONGATION FACTOR G (E	4.55e+00
15	66	28.5	118	2	024774	REPON RNA POLYMERASE SI	6.73e+00
16	66	29.5	134	10	09X9K6	RIBOSOMAL PROTEIN S14.	6.73e+00
17	66	28.5	234	2	Q55877	HYPOTHETICAL 25.9 KD P	6.73e+00
18	66	29.5	308	14	098494	GENOME, PARTIAL SEQUEN	6.73e+00
19	66	29.5	320	5	P91523	COSMID T28A11.	6.73e+00
20	66	29.5	442	2	069406	BETAINE REDUCTASE.	6.73e+00

DR EMBL; Y1

DR EMBL; Y1


```

Best Local Similarity 45.58; Pred.No. 6.07e-01;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 129 FIGLIELGFAQIKEELEDCEA 150
    |||: ||| :|| | :| :|
QY 9 FIGITELGGHEIKVLVPGCHG 30
    |||: ||| :|| | :| :|

RESULT 5
ID ID Q9YLX6 PRELIMINARY; PRT; 471 AA.
AC Q9YLX6;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE SPTP4 (EC 3.1.3.48) (FRAGMENT).
OS Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haplosclerida; Spongillidae; Ephydatia.
[1]
RN RN SEQUENCE FROM N.A.
RP RP MEDLINE; 99246376.
RX RA ONO K., SUGA H., IWABE N., KUMA K., MIYATA T.;
RT "Multiple protein tyrosine phosphatases in sponges and explosive gene
RT duplication in the early evolution of animals before the parazoan-
RT eumetazoan split.";
RL J. Mol. Evol. 48:654-662(1999).
DR EMBL; AB019125; BAA82558.1;
DR PROSITE; PS00363; TYR_PHOSPHATASE_1; 2.
KW Hydrolase.
KW NON_TER
FT FT 1
SQ SEQUENCE 471 AA; 54509 MW; BA496490 CRC32;

Query Match 32.18; Score 72; DB 5; Length 471;
Best Local Similarity 22.68; Pred.No. 6.07e-01;
Matches 7; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

Db 223 HYNKKNRYTNIVPFDHSRCSRIIPGVEGS 253
    ||: ::::: | :| :| :| :| :|
QY 1 QYIKANSKFIGITELGGHEIKVLVPGCHGS 31
    ||: ::::: | :| :| :| :| :|

RESULT 6
ID ID Q22814 PRELIMINARY; PRT; 208 AA.
AC Q22814;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE T26C5.1 PROTEIN.
DE DE T26C5.1 PROTEIN.
GN GN T26C5.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RN RN SEQUENCE FROM N.A.
RP RP THOMAS K.;
RL RA THOMAS K.;
RN RN 2
SQ SEQUENCE FROM N.A.
RP RP MEDLINE; 94150718.
RX RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z50859; CAA90726.1;

```

SQ SEQUENCE 841 AA; 96367 MW; E7C2968A CRC32;
 Query Match 31.3%; Score 70; DB 10; Length 841;
 Best Local Similarity 42.1%; Pred. No. 1.37e+00;
 Matches 8; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
 Db 264 TKVAFAEALGDNEVEKALV 282
 :|:|||||:|:|
 QY 7 SKFIGITELGGHEIKKVLV 25
 :|:|||||:|:|
 RESULT 9
 ID Q9ZBM6 PRELIMINARY; PRT; 189 AA.
 AC Q9ZBM6;
 DT 01-WAY-1999 (TrEMBLrel. 10, Created)
 DT 01-WAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-WAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.
 GS MLCB1450.06C.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BROWN D., CHURCHER C.M.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA JAMES K.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EIGLMEIER K., HONORE N., WOODS S.A., CAUDRON B., COLE S.T.;
 "Use of an ordered cosmid library to deduce the genomic organization
 of Mycobacterium leprae";
 RL Mol. Microbiol. 7:197-206(1993).
 DR EMBL; AL035159; CAA22690.1; -.
 SQ SEQUENCE 189 AA; 19894 MW; 42073499 CRC32;
 Query Match 30.4%; Score 68; DB 2; Length 189;
 Best Local Similarity 44.4%; Pred. No. 3.07e+00;
 Matches 8; Conservative 7; Mismatches 2; Indels 1; Gaps 1;
 Db 149 TDGVREKXSVLY-GCOA 165
 :|:|:|:|:|:|:|
 QY 13 TELGGHEIKKVLVPGCHG 30
 :|:|:|:|:|:|:|
 RESULT 10
 ID P89671 PRELIMINARY; PRT; 224 AA.
 AC P89671;
 DT 01-WAY-1997 (TrEMBLrel. 03, Created)
 DT 01-WAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
 DE CAPSID PROTEIN (FRAGMENT).
 OS San Miguel sea lion virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
 OC Calicivirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NEILL J.D., MEYER R.F., SEAL B.S.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U76886; AAB48432.1; -.
 FT NON_TER 1
 NON_TER 224 224
 SQ SEQUENCE 224 AA; 24382 MW; 75AA14BA CRC32;
 Query Match 30.4%; Score 68; DB 14; Length 224;
 Best Local Similarity 38.9%; Pred. No. 3.07e+00;
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
 Db 115 KSO5VYITADLGNTVK 132


```

RESULT 15
ID Q24774; PRELIMINARY; PRT; 118 AA.
AC Q24774;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE RPN RNA POLYMERASE SIGMA FACTOR N, PARTIAL AND COMPLETE CDS
DE (FRAGMENT).
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=138-2;
RX MEDLINE; 98012988.
RA KAWAGISHI I., NAKADA M., NISHIOKA N., HOMMA M.;
RT "Cloning of a Vibrio alginolyticus rpn gene that is required for
RT polar flagellar formation.";
RL J. Bacteriol. 179:6851-6854(1997).
DR EMBL; AB006709; BAA21882.1;
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
FT NON_TER 1
SQ SEQUENCE 118 AA; 13153 MW; 39E91291 CRC32;

Query Match 29.5%; Score 66; DB 2; Length 118;
Best Local Similarity 29.2%; Pred. No. 6.73e+00;
Matches 7; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

Db 34 QFILLDEPAGVDPIVNDIKKII 57
QY 1 QYKANSKFIGITELGGHEIKKVL 24

```

Search completed: Sun Apr 2 17:49:11 2000
 Job time : 73 secs.

RELEASE

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:47:09 2000; Maspar time 28.18 Seconds
Tabular output not generated. 32.853 Million cell updates/sec

Title: >US-09-362-731-1
Description: (1-31) from US09362731.pep
Perfect Score: 224
Sequence: 1 QYKANSKFIGITELGGHEIKKVLVPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:Swissprot

Statistics: Mean 32.909; Variance 46.340; scale 0.710

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	113	50.4	146	1	DER2_DERPT MITE ALLERGEN DER P 2	1.79e-09
2	104	46.4	1314	1	TETX_CLOTE TETANUS TOXIN PRECURSO	1.72e-07
3	92	41.1	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	5.89e-05
4	74	33.0	417	1	DHML_PARDE METHYLAMINE DEHYDROGEN	1.82e-01
5	71	31.7	382	1	TGFI_XENLA TRANSFORMING GROWTH FA	6.26e-01
6	69	30.8	287	1	TRUE_AQUAE TRNA PSEUDOURIDINE SYN	1.40e+00
7	69	30.8	364	1	DP3B_MYCPN DNA POLYMERASE III, BE	1.40e+00
8	69	30.8	380	1	DP3B_MYCPN DNA POLYMERASE III, BE	1.40e+00
9	69	30.8	618	1	NOP2_YEAST NUCLEOLAR PROTEIN NOP2	1.40e+00
10	69	30.8	691	1	EFG_HELPY ELONGATION FACTOR G (E	1.40e+00
11	68	30.4	1882	1	Y468_MYCPN HYPOTHETICAL PROTEIN M	2.07e+00
12	67	29.9	310	1	PYRB_SALTY ASPARTATE CARBAMOYLTRA	3.06e+00
13	67	29.9	485	1	HXBK_YEAST HEXOKINASE B (EC 2.7.1	3.06e+00
14	66	29.5	140	1	YB17_YEAST HYPOTHETICAL 15.8 KD P	4.50e+00
15	66	29.5	188	1	LCAT_PIG PHOSPHATIDYLCHOLINE-ST	4.50e+00
16	66	29.5	213	1	KAD_MYCCA ADENYLATE KINASE (EC 2	4.50e+00
17	66	29.5	438	1	LCAT_MOUSE PHOSPHATIDYLCHOLINE-ST	4.50e+00
18	66	29.5	440	1	LCAT_PAPAN PHOSPHATIDYLCHOLINE-ST	4.50e+00
19	66	29.5	440	1	LCAT_HUMAN PHOSPHATIDYLCHOLINE-ST	4.50e+00
20	66	29.5	440	1	LCAT_RABIT PHOSPHATIDYLCHOLINE-ST	4.50e+00
21	66	29.5	808	1	SYFE_SYN7 PHENYLALANYL-TRNA SYN	4.50e+00
22	66	29.5	1035	1	POLY_DROME RETROVIRUS-RELATED POL	4.50e+00
23	65	29.0	129	1	EXBD_HAEDU BIOPOLYMER TRANSPORT E	6.58e+00

24	65	29.0	145	1	EXBD_PASHA BIOPOLYMER TRANSPORT E	6.58e+00
25	65	29.0	207	1	REGQ_LAMBD ANTITERMINATION PROTEIN	6.58e+00
26	65	29.0	221	1	Y805_METUA HYPOTHETICAL PROTEIN M	6.58e+00
27	65	29.0	226	1	TRPF_METUA N-(5'-PHOSPHORIBOSYL)A	6.58e+00
28	65	29.0	305	1	PYRB_SERWA ASPARTATE CARBAMOYLTRA	6.58e+00
29	65	29.0	339	1	GPDA_ECOLI GLYCEROL-3-PHOSPHATE D	6.58e+00
30	65	29.0	379	1	HYPD_ALCEU HYDROGENASE EXPRESSION	6.58e+00
31	65	29.0	440	1	LCAT_RAT PHOSPHATIDYLCHOLINE-ST	6.58e+00
32	65	29.0	455	1	RADA_TREPA DNA REPAIR PROTEIN RAD	6.58e+00
33	65	29.0	485	1	HXKA_YEAST HEXOKINASE A (EC 2.7.1	6.58e+00
34	65	29.0	487	1	HUS2_YEAST HUS2 PROTEIN.	6.58e+00
35	65	29.0	614	1	DNAK_ODOSI DNAK PROTEIN (HEAT SHO	6.58e+00
36	64	28.6	151	1	RS14_DROME 40S RIBOSOMAL PROTEIN	9.59e+00
37	64	28.6	171	1	GPDA_SALTY GLYCEROL-3-PHOSPHATE D	9.59e+00
38	64	28.6	329	1	NADA_CYPAA QUINOLINATE SYNTHETASE	9.59e+00
39	64	28.6	396	1	YCGA_BACSU HYPOTHETICAL 41.5 KD P	9.59e+00
40	64	28.6	415	1	AMSJ_ERWAM AMYLOVORAN BIOSYNTHESI	9.59e+00
41	64	28.6	425	1	DHMH_PARVE METHYLAMINE DEHYDROGEN	9.59e+00
42	64	28.6	451	1	MURD_BACSU UDP-N-ACETYLURAMIDYLAL	9.59e+00
43	64	28.6	590	1	2AAA_DROME PROTEIN PHOSPHATASE PP	9.59e+00
44	64	28.6	631	1	XRC1_MOUSE DNA-REPAIR PROTEIN XRC	9.59e+00
45	64	28.6	633	1	XRC1_HUMAN DNA-REPAIR PROTEIN XRC	9.59e+00

ALIGNMENTS

RESULT 1
ID DER2_DERPT STANDARD; PRT; 146 AA.
AC P49278;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
GN DERP2.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90256301.
RA CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
RA THOMAS W.R.;
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
RT plaque immunoassay.";
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
RN [2]
RP PARTIAL SEQUENCE OF 18-57.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATT'S-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE; 98409423.
RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
RT "Tertiary structure of the major house dust mite allergen Der p 2:
RT sequential and structural homologies.";
RL Biochemistry 37:12707-12714(1998).
CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
DR PDB; 1A9V; 14-OCT-98.
KW Allergen; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER P 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;

Query Match 50.4%; Score 113; DB 1; Length 146;
Best Local Similarity 63.6%; Pred. No. 1.79e-09;

Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 20 VDVKCANHEIKKVLPGCHGS 41
 QY 10 IGITELGGHEIKKVLPGCHGS 31

RESULT 2
 ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: primary structure, expression in E. coli, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911.
 RX MEDLINE; 87040747.
 RA FAIRWEATHER N.F., LYNESS V.A.;
 RL "The complete nucleotide sequence of tetanus toxin.";
 RN Nucleic Acids Res. 14:7809-7812(1986).
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE; 86085672.
 RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in Escherichia coli.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE; 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulphydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE; 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]
 RP IDENTIFICATION AS ZINC-PROTEASE.
 RX MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RT neurotransmitter release and protease activity depend on zinc.";
 RL EMBO J. 11:3577-3583(1992).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL Nature 359:832-835(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.

RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL Nat. Struct. Biol. 4:788-792(1997).
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSOMAL-2
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSOMAL-2
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
 CC GANGLIOSIDE RECEPTORS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X04436; CAA28033.1; -;
 CC EMBL; M12739; AAA3282.1; -;
 CC EMBL; X06214; CAA29564.1; -;
 CC PIR; A25689; BTCLFN.
 CC PDB; 1AF9; 29-APR-98.
 CC PDB; 1ABD; 14-OCT-98.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
 CC 3D-structure.
 CC INIT_MET 0
 CC CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 CC CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 CC METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACT_SITE 233 233 BY SIMILARITY.
 CC METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 CC TRANSMEM 226 246 POTENTIAL.
 CC TRANSMEM 669 689 POTENTIAL.
 CC DISULFID 438 466 INTERCHAIN.
 CC DISULFID 1076 1092
 CC SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
 SQ

Query Match 46.48; Score 104; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 1.72e-07;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYIKANSKFIGITEL 843
 QY 1 QYIKANSKFIGITEL 15

RESULT 3
 ID DEF2_DERFA STANDARD; PRT; 146 AA.
 AC Q00835; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERF2
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
 OC Dermatophagoides.

Query Match 41.1%; Score 92; DB 1; Length 146;
Best Local Similarity 50.0%; Pred. No. 5,89e-05;
Matches 11; Conservative 6; Mismatches 5; Indels

[illegible]

DE SYNTHASE) (PSI55 SYNTHASE) (PSEUDOURIDYLATE SYNTHASE) (URACIL
DE HYDROLYASE).
GN TRUB OR AQ_705.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.,
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -1- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: URACIL + D-RIBOSE 5-PHOSPHATE = PSEUDOURIDINE
CC 5'-PHOSPHATE + H(2)O.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000703; AAC06885.1; -
DR PFAM; PF01509; TRUB_N.1.
KW Lyase; tRNA processing.
SQ SEQUENCE 287 AA; 32259 MW; 633369D0 CRC32;

Query Match 30.8%; Score 69; DB 1; Length 287;
Best Local Similarity 59.18; Pred. No. 1.40e+00;
Matches 13; Conservative 3; Mismatches 4; Indels 2; Gaps 2;

Db 266 DSKFIGIGELKGGVLSPKRLLV 287
:||||| || || ||:|
QY 6 NSKFIGITEL-GGH-EIKKVLV 25

RESULT 7
ID DP3B_MYCGE STANDARD; PRT; 364 AA.
AC P47247;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7).
GN DNAN OR MG001.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 96026346.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,
RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.,
RT "The minimal gene complement of Mycoplasma genitalium".
RL Science 270:397-403(1995).
RN [2]
RN REVISIONS.
RA FRASER C.M., GOCAYNE J.D., WHITE O., ADAMS M.D., CLAYTON R.A.,
RA FLEISCHMANN R.D., BULT C.J., KERLAVAGE A.R., SUTTON G., KELLEY J.M.,
RA FRITZCHMAN J.L., WEIDMAN J.F., SMALL K.V., SANDUSKY M., FUHRMANN J.L.,
RA NGUYEN D.T., UTTERBACK T.R., SAUDEK D.M., PHILLIPS C.A., MERRICK J.M.,
RA TOMB J.-F., DOUGHERTY B.A., BOTT K.F., HU P.-C., LUCIER T.S.,

RA PETERSON S.N., SMITH H.O., HUTCHISON C.A. III, VENTER J.C.;
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE OF 267-364 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE; 94364962.
RA BAILEY C.C., BOTT K.F.;
RT "An unusual gene containing a dnaJ N-terminal box flanks the putative
RT origin of replication of Mycoplasma genitalium".
RL J. Bacteriol. 176:5814-5819(1994).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION AND ATP-
CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-
CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND
CC THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE
CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-
CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,
CC GAMMA, AND DELTA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39679; AAC71217.1; -
DR EMBL; U09251; FAA57069.1; -
DR TIGR; MG001; -
DR PFAM; PF00712; DNA_pol3_beta; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
SQ SEQUENCE 364 AA; 42399 MW; 0CE3F6A2 CRC32;

Query Match 30.8%; Score 69; DB 1; Length 364;
Best Local Similarity 40.0%; Pred. No. 1.40e+00;
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 339 FQNGKYFLITSKEPELKLVLPS 363
:||||| || || ||:|
QY 3 IKANSKFIGITELGGHEIKKVLVPG 27

RESULT 8
ID DP3B_MYCPN STANDARD; PRT; 380 AA.
AC Q50313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE III, BETA CHAIN (EC 2.7.7.7).
GN DNAN.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 96177562.
RA HUBERT H., HIMMELREICH R., FLAGENS H., HERRMANN R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
RT Mycoplasma pneumoniae comprising the dnaA region, the atp operon and
RT a cluster of ribosomal protein genes".
RL Nucleic Acids Res. 24:628-639(1996).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 97105885.

RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
 RA HERRMANN R.;
 RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*
 RL *pneumoniae*.";
 CC Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE BETA CHAIN IS REQUIRED FOR INITIATION OF REPLICATION ONCE IT
 CC IS CLAMPED ONTO DNA, IT SLIDES FREELY (BIDIRECTIONAL AND ATP-
 CC INDEPENDENT) ALONG DUPLEX DNA (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N PROPHOSPHATE + DNA(N).
 CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND
 CC THETA CHAINS) THAT CAN REPAIR SHORT GAPS CREATED BY NUCLEASE
 CC IN DUPLEX DNA. FOR EFFICIENT REPLICATION OF THE LONG, SINGLE-
 CC STRANDED TEMPLATES, POL III REQUIRES THE AUXILIARY CHAINS BETA,
 CC GAMMA, AND DELTA (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U34816; AAC3645.1; -
 CC EMBL; AB000017; AB95801.1; -
 CC DR PFAM; PF00712; DNA_pol3_beta.1.
 CC KW Transferase; DNA-directed DNA polymerase; DNA replication.
 CC SQ SEQUENCE 380 AA; 43856 MW; 8464DBD6 CRC32;
 Query Match 30.8%; Score 69; DB 1; Length 380;
 Best Local Similarity 30.8%; Pred. No. 1.40e+00;
 Matches 8; Conservative 10; Mismatches 8; Indels 0; Gaps 0;
 Db 354 YFQSGNRYFLISSNPELKEILVPS 379
 QY 2 YIKANSFEIGITELGGHEIKKVLPG 27
 RESULT 9
 ID N0P2_YEAST STANDARD; PRT; 618 AA.
 AC P40991;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1999 (Rel. 39, Last annotation update)
 DE NUCLEOLAR PROTEIN N0P2.
 GN N0P2 OR YNA1 OR YNL061W OR N2428 OR YNL2428W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BJ2168;
 RX MEDLINE; 95105226.
 RA DE BEUS E., BROCKENBROUGH J.S., HONG B., ARIS J.P.;
 RT "Yeast N0P2 encodes an essential nucleolar protein with homology to a
 RT human proliferation marker.";
 RL J. Cell Biol. 127:1799-1813(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1676;
 RX MEDLINE; 96021608.
 RA BERGEZ P., DOIGNON F., CROUZET M.;
 RT "The sequence of a 44 420 bp fragment located on the left arm of
 RT chromosome XIV from *Saccharomyces cerevisiae*.";
 RL Yeast 11:967-974(1995).
 RN [3]
 RP ERRATUM.
 RX MEDLINE; 97060022.

RA BERGEZ P., DOIGNON F., CROUZET M.;
 RL Yeast 12:297-297(1996).
 RN [4]
 RC SEQUENCE OF 146-618 FROM N.A.
 RP STRAIN-S288C;
 RA GARCIA-BARRIO M.T., CUESTA R., HINNEBUSCH A.G., TAMAME GONZALEZ M.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD BE INVOLVED IN NUCLEOLAR FUNCTION DURING THE ONSET
 CC OF GROWTH, AND IN THE MAINTENANCE OF NUCLEOLAR STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
 CC -1- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
 CC NOLL/NOP2 (EUKARYOTES) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; X82656; CAA57979.1; -
 CC EMBL; U12141; AAA9650.1; -
 CC DR EMBL; Z71337; CAA95934.1; -
 CC DR EMBL; X83512; CAA58502.1; -
 CC DR SGD; L0002535; YNA1.
 CC DR PROSITE; PS01153; NOLL_NOP2_SUN; 1.
 CC DR PFAM; PF01189; Noll_Nop2_Sun; 1.
 CC KW Nuclear protein; Ribosome biogenesis.
 CC FT CONFLICT 577 577 I -> M (IN REF. 2).
 CC SQ SEQUENCE 618 AA; 69812 MW; 0EA01260 CRC32;
 Query Match 30.8%; Score 69; DB 1; Length 618;
 Best Local Similarity 36.0%; Pred. No. 1.40e+00;
 Matches 9; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
 Db 326 HYIQAASSFLPVLDPHENERIL 350
 QY 1 QYI-KANSKFIGITELGGHEIKKVL 24
 RESULT 10
 ID EFG_HELPY STANDARD; PRT; 691 AA.
 AC P56002;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE ELONGATION FACTOR G (EF-G).
 GN FUSA OR HP1195.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-26695 / ATCC 700392;
 RX MEDLINE; 97394467.
 RA TOMB J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.;
 RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
 RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
 RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
 RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
 RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
 RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
 RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
 RA VENTER J.C.;
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 CC RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

OY 1 QYIKANSKFIGITEL-GGHEIKKVLVP 26

RESULT 13
ID HXKB YEAST STANDARD; PRT; 485 AA.
AC P04807;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DT 15-DEC-1999 (Rel. 39, Last annotation update)
DE HEXOKINASE B (EC 2.7.1.1) (HEXOKINASE PII).
GN HXK2 OR HKB OR HEX1 OR YGL253W OR NRB486.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 86120382.
RA STACHELEK C., STACHELEK J., SWAN J., BOTSTEIN D., KONIGSBERG W.;
RT "Identification, cloning and sequence determination of the genes
RL specifying hexokinase A and B from yeast.";
RL Nucleic Acids Res. 14:945-963(1986).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE: 86056943.
RA FROEHLICH K.-U., ENTIAN K.-D., MECKE D.;
RT "The primary structure of the yeast hexokinase PII gene (HXK2) which
RL is responsible for glucose repression.";
RL Gene 36:105-111(1985).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN-S288C / FY1679;
RX MEDLINE: 97127827.
RA COISSAC E., MAILLIER E., ROBINEAU S., NETTER P.;
RT "Sequence of a 39,411 bp DNA fragment covering the left end of
RL chromosome VII of Saccharomyces cerevisiae.";
RL yeast 12:1555-1562(1996).
[4]
RN SEQUENCE OF 1-246 FROM N.A.
RC STRAIN-W303;
RX MEDLINE: 93311123.
RA BREITWIESER W., PRICE C., SCHUSTER T.;
RT "Identification of a gene encoding a novel zinc finger protein in
RL Saccharomyces cerevisiae.";
RL yeast 9:551-556(1993).
[5]
RN SEQUENCE OF 118-126: 175-184 AND 303-313.
RC STRAIN-ATCC 38531 / Y41;
RX MEDLINE: 95255188.
RA NORBECK J., BLOMBERG A.;
RT "Gene linkage of two-dimensional polyacrylamide gel electrophoresis
RL resolved proteins from isogene families in Saccharomyces cerevisiae
RL by microsequencing of in-gel trypsin generated peptides.";
RL Electrophoresis 16:149-156(1995).
[6]
RN PHOSPHORYLATION OF SER-14.
RX MEDLINE: 94114477.
RA KRIEDEL T.M., RUSH J., VOJTEK A.B., CLIFTON D., FRAENKEL D.G.;
RT "In vivo phosphorylation site of hexokinase 2 in Saccharomyces
RL cerevisiae.";
RL Biochemistry 33:148-152(1994).
[7]
RN PHOSPHORYLATION OF SER-157.
RX MEDLINE: 97199316.
RA HEIDRICH K., OTTO A., BEHLKE J., RUSH J., WENZEL K.W., KRIEDEL T.;
RT "Autophosphorylation-inactivation site of hexokinase 2 in
RL Saccharomyces cerevisiae.";
RL Biochemistry 36:1960-1964(1997).
[8]
RN SEQUENCE OF 1-18, AND PHOSPHORYLATION OF SER-14.
RX MEDLINE: 98384167.
RA BEHLKE J., HEIDRICH K., NAUMANN M., MULLER E.-C., OTTO A., REUTER R.,
RA KRIEDEL T.;
RT "Hexokinase 2 from Saccharomyces cerevisiae: regulation of oligomeric

RT structure by in vivo phosphorylation at serine-14.";
RL Biochemistry 37:11989-11995(1998).
[9]
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE: 78244654.
RA ANDERSON C.M., STENKAMP R.E., STEITZ T.A.;
RT "Sequencing a protein by X-ray crystallography. II. Refinement of
RT yeast hexokinase B co-ordinates and sequence at 2.1-A resolution.";
J. Mol. Biol. 123:15-33(1978).
CC -1- FUNCTION: MAIN GLUCOSE PHOSPHORYLATING ENZYME. MAY PLAY A
CC REGULATORY ROLE IN BOTH INDUCTION AND REPRESSION OF GENE
CC EXPRESSION BY GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ATP + D-HEXOSE -> ADP + D-HEXOSE 6-PHOSPHATE.
CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
CC INHIBITION BY ATP.
CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/H/HK.html".

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

DR EMBL: X03483; CAA27203.1; -
DR EMBL: M11181; AAA34697.1; -
DR EMBL: X94357; CAA64134.1; -
DR EMBL: 272775; CAA96973.1; -
DR EMBL: X67787; CAA48003.1; -
DR PIR: B23523; KIBYHB.
DR PIR: S28555; S28555.
DR PDB: 2YHX; 15-JUL-92.
DR SWISS-2DPAGE: P04807; YEAST.
DR YEPD: 8536; -
DR YEPD: 8548; -
DR SGD: L0000834; HXK2.
DR PROSITE: PS00378; HEXOKINASES; 1.
DR PFAM: PF00349; hexokinase; 1.
KW Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW 3D-structure; Phosphorylation.
FT INIT_MET 0
FT BINDING 110 110 ATP (BY SIMILARITY).
FT DOMAIN 151 177 GLUCOSE-BINDING (POTENTIAL).
FT MOD_RES 14 14 PHOSPHORYLATION.
FT MOD_RES 157 157 PHOSPHORYLATION.
FT CONFLICT 28 28 N -> I (IN REF. 1).
FT CONFLICT 32 32 I -> N (IN REF. 2).
FT CONFLICT 60 60 G -> V (IN REF. 1).
FT CONFLICT 196 196 T -> S (IN REF. 1).
FT CONFLICT 420 421 YN -> ST (IN REF. 2).
FT CONFLICT 443 444 TS -> PH (IN REF. 2).
FT CONFLICT 452 452 I -> V (IN REF. 2).
FT CONFLICT 461 461 A -> P (IN REF. 2).
FT HELIX 23 33
FT HELIX 37 55
FT STRAND 65 66
FT STRAND 79 86
FT STRAND 90 99
FT TURN 100 101
FT STRAND 102 110
FT STRAND 112 112
FT TURN 115 118
FT TURN 123 123
FT HELIX 124 141
FT TURN 142 142
FT STRAND 150 155

FT STRAND 164 164
 FT TURN 165 166
 FT STRAND 169 169
 FT TURN 175 176
 FT STRAND 187 187
 FT STRAND 188 189
 FT STRAND 202 208
 FT STRAND 210 221
 FT STRAND 220 224
 FT STRAND 225 231
 FT STRAND 235 241
 FT STRAND 243 245
 FT TURN 251 252
 FT STRAND 263 266
 FT TURN 270 275
 FT STRAND 283 291
 FT STRAND 299 304
 FT STRAND 306 322
 FT TURN 323 324
 FT TURN 338 339
 FT TURN 343 343
 FT STRAND 344 351
 FT STRAND 358 368
 FT TURN 369 369
 FT STRAND 374 395
 FT TURN 396 396
 FT STRAND 397 406
 FT TURN 407 407
 FT STRAND 411 416
 FT TURN 418 421
 FT TURN 424 425
 FT STRAND 426 438
 FT STRAND 445 447
 FT TURN 450 454
 FT TURN 458 460
 FT STRAND 461 475
 FT STRAND 485 AA; 53811 MW; 6C3BE887 CRC32;

Query Match 29.9%; Score 67; DB 1; Length 485;
 Best Local Similarity 40.9%; Pred. No. 3.06e+00;
 Matches 9; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 76 KESGDFLAI-DLGGTNRVLV 96
 QY 4 KANSKFIGITELGGHKKVLV 25

RESULT 14
 ID YB17 YEAST STANDARD; PRT; 140 AA.
 AC P38291;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE HYPOTHETICAL 15.8 KD PROTEIN IN TYR1-SSE2 INTERGENIC REGION.
 GN YBR167C OR YBR1219.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA ENTIAN K.-D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,
 RA NIEGEMANN E., SCHENK-GROENINGER R., SERVOS J., WEHNER E.,
 RA WOLTER R., BRENDEN M., BAUER J., BRAUN H., DERN K., DUESTERHUS S.,
 RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREMS B., PROFT M.,
 RA SIEGERS K., BAUR A., BOLES E., MIOSGA T.,
 RA SCHAAFF-GERSTENSCHLAGER I., ZIMMERMANN F.K.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z36036; CAA85128.1; --
 DR PIR; S46038; S46038.
 KW Hypothetical protein.
 SQ SEQUENCE 140 AA; 15814 MW; 6A571113 CRC32;

Query Match 29.5%; Score 66; DB 1; Length 140;
 Best Local Similarity 30.8%; Pred. No. 4.50e+00;
 Matches 8; Conservative 6; Mismatches 11; Indels 1; Gaps 1;

Db 57 KQSSYVAVLGMG-KAVEKTIALGCH 81
 QY 4 KANSKFIGITELGGHKKVLVPGCH 29

RESULT 15
 ID LCAT_PIG STANDARD; PRT; 188 AA.
 AC P30930;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PHOSPHATIDYLCHOLINE-STEROL ACYLTRANSFERASE (EC 2.3.1.43) (LECITHIN-
 DE CHOLESTEROL ACYLTRANSFERASE) (PHOSPHOLIPID-CHOLESTEROL
 DE ACYLTRANSFERASE) (FRAGMENTS).
 GN LCAT.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-PLASMA;
 RX MEDLINE; 90075705.
 RA YUEKSEL K.U., PARK Y.B., JUNG J., GRACY R.W., LACKO A.G.;
 RT "Studies on the structure of lecithin:cholesterol acyltransferase
 RT (LACAT) -- comparisons of the active site region and secondary
 RT structure of the human and the porcine enzymes.";
 RL Comp. Biochem. Physiol. 94B:389-394(1989).
 RN [2]
 RP SEQUENCE OF 97-106 AND 142-154.
 RC TISSUE-PLASMA;
 RX MEDLINE; 87156771.

RA PARK Y.B., YUEKSEL K.U., GRACY R.W., LACKO A.G.;
 RT "The catalytic center of lecithin:cholesterol acyltransferase:
 RT Isolation and sequence of diisopropyl fluorophosphate-labeled
 RT peptides.";
 RL Biochem. Biophys. Res. Commun. 143:360-363(1987).
 CC -1- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
 CC ACT AS ACCEPTOR).
 CC -1- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
 CC THIS ENZYME.
 CC -1- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC PIR; P0153; P0153.
 DR PIR; A29544; A29544.
 DR PIR; B29544; B29544.
 DR PROSITE; PS00120; LIPASE_SER; PARTIAL.
 KW Transferase: Acyltransferase; Lipid metabolism; Glycoprotein.
 FT CARBOHYD 20 20
 FT NON_CONS 34 35
 FT NON_CONS 44 45
 FT NON_CONS 60 61
 FT NON_CONS 66 67
 FT NON_CONS 77 78
 FT NON_CONS 84 85
 FT NON_CONS 96 97

FT NON_CONS 106 107
FT NON_CONS 115 116
FT NON_CONS 154 135
FT NON_TER 188 188
SQ SEQUENCE 188 AA; 21232 MW; 876EB48E CRC32;

Query Match 29.5%; Score 66; DB 1; Length 188;
Best Local Similarity 47.4%; Pred. No. 4.50e+00;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Db 16 AELSNHTRPVILVPGCLGN 34
QY 13 TELGGHEIKKVLVPGCHGS 31

Search completed: Sun Apr 2 17:47:40 2000
Job time : 31 secs.

(TM)

Result No.	Query		Length	DB	ID	Description	Pred. No.
	Score	Match					
1	113	50.4	145	2	A60381	allergen Der p II pre	7.50e-08
2	104	46.4	1315	1	B7CLTN	tentoxylsin (EC 3.4.	4.16e-06
3	92	41.1	129	2	A61501	allergen Der f II - h	7.10e-04
4	92	41.1	129	2	JU0394	allergen Der f II (pF	7.10e-04
5	92	41.1	138	2	A61241	allergen Der f II pre	7.10e-04
6	92	41.1	138	2	B61241	allergen Der f II pre	7.10e-04
7	74	33.0	417	2	JU0660	amine dehydrogenase (8.72e-01
8	73	32.6	205	2	F71663	sco2 protein precursor	1.26e+00
9	71	31.7	382	2	B61036	transforming growth f	2.61e+00
10	70	31.3	393	2	A71201	hypothetical protein	3.74e+00
11	69	30.8	267	2	A64200	DNA polymerase III be	5.33e+00
12	69	30.8	287	2	F70361	RNA-pseudouridine sy	5.33e+00
13	69	30.8	380	2	S62836	DNA-directed DNA poly	5.33e+00
14	69	30.8	618	2	A55188	nucleolar protein NOP	5.33e+00
15	69	30.8	692	2	C64669	translation elongatio	5.33e+00
16	68	30.4	1882	2	S73484	hypothetical protein	7.58e+00
17	67	29.9	138	2	H75114	co-induced hydrogenas	1.07e+01
18	67	29.9	311	1	OWEBAC	aspartate carbamoylitr	1.07e+01
19	67	29.9	486	1	KIBYHB	hexokinase (EC 2.7.1.	1.07e+01
20	67	29.9	517	2	S62667	Nramp1 protein - rice	1.07e+01
21	67	29.9	692	2	G71847	elongation factor 9 (1.07e+01
22	66	29.5	140	2	S46038	hypothetical protein	1.51e+01
23	66	29.5	188	2	PL0153	phosphatidylcholine--	1.51e+01

homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
#molecule_type DNA
#residues 1-1315 #label EIS
#cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
#molecule_type DNA
#residues 1-1315 #label FAI
#cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
#experimental_source strain CN3911
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.
#journal J. Bacteriol. (1986) 165:21-27
#title Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.
#cross-references MUID:86085672
#accession A25194
#molecule_type DNA
#residues 743-1315 #label FA2
#cross-references GB:M12739; NID:g144920; PIDN:AAA23282.1; PID:g144921
#accession B25194
#molecule_type protein
#residues 865-894 #label FA3
REFERENCE A60759
#authors Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
#journal Infect. Immun. (1989) 57:3588-3593
#title Isolation, purification, and characterization of fragment B, the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 #label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell epitopes.
#cross-references MUID:89093918
#contents annotation; epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
#molecule_type protein
#residues 2-31 #label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B,C) chains that are covalently linked by an interchain disulfide bond (the individual chains are not toxic when separated). The amino end of the heavy chain (fragment B) can be separated from the carboxyl end (fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds

to gangliosides and may target the toxin to the motor end plate. Fragment A is a zinc-dependent endopeptidase. This potent neurotoxin binds to peripheral neuronal synapses, is internalized, and moves by retrograde transport up the axon into the spinal cord where it can move between postsynaptic and presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin (vesicle-associated membrane protein 2).

FUNCTION #description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
FEATURE 2-457 : #product tentoxylisin light chain (fragment A) #status predicted #label TRU
461-1315 : #product tentoxylisin heavy chain (fragment B,C) #status predicted #label TRU
461-864 : #domain channel forming (fragment B) #status predicted #label TRU
865-1315 : #domain ganglioside binding (fragment C) #status predicted #label TRU
233,237 : #binding_site zinc (His) #status predicted
234 : #active_site Glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 46.4% Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 4.16e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 QYIKANSKFGITEL 844
|||||
QY 1 QYIKANSKFGITEL 15

RESULT 3
ENTRY #type fragment
TITLE allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
ACCESSIONS A61501
REFERENCE A61501
#authors Trudinger, M.; Chua, K.Y.; Thomas, W.R.
#journal Clin. Exp. Allergy (1991) 21:33-37
#title CDNA encoding the major mite allergen Der f II.
#cross-references MUID:91215495
#accession A61501
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-129 #label TRU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476
Query Match 41.1% Score 92; DB 2; Length 129;
Best Local Similarity 50.0%; Pred. No. 7.10e-04;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 3 VDVKDCANNEIKKVMVDGCHGS 24
: : : : :
QY 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 4
ENTRY #type complete
TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
ACCESSIONS JU0394


```

REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
#journal Haida, M.; Okudaira, H.
#title Agric. Biol. Chem. (1991) 55:1233-1238
#cross-references MIM:91291341
#accession J00394
##molecule_type mRNA
##residues 1-129 ##label YUU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 41.1%; Score 92; DB 2; Length 129;
Best Local Similarity 50.0%; Pred. No. 7.10e-04;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 3 DVVKDCANNEIKKVMVDCGCHGS 24
QY 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 5
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS A61241; PS0417
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
#journal Haida, M.; Dohi, M.; Okudaira, H.
#title Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#cross-references MIM:92040281
#accession A61241
##molecule_type mRNA
##residues 1-138 ##label YUU
##note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing

CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status experimental #label
MAT
#length 138 #checksum 2894

Query Match 41.1%; Score 92; DB 2; Length 138;
Best Local Similarity 50.0%; Pred. No. 7.10e-04;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 12 DVVKDCANNEIKKVMVDCGCHGS 33
QY 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 6
ENTRY B61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS B61241; J00395
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
#journal Haida, M.; Dohi, M.; Okudaira, H.
#title Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#cross-references MIM:92040281
#accession B61241
##molecule_type mRNA
##residues 1-138 ##label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status experimental #label
MAT
#length 138 #checksum 2894

Query Match 41.1%; Score 92; DB 2; Length 138;
Best Local Similarity 50.0%; Pred. No. 7.10e-04;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 12 DVVKDCANNEIKKVMVDCGCHGS 33
QY 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 7
ENTRY JH0660 #type complete
TITLE amine dehydrogenase (EC 1.4.99.3) large chain precursor -
Paracoccus denitrificans
ALTERNATE_NAMES methylamine dehydrogenase large chain
ORGANISM #formal_name Paracoccus denitrificans
DATE 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change
23-Jul-1999
ACCESSIONS JH0660; S51049
REFERENCE PH0856
#authors Chistoserdov, A.Y.; Boyd, J.; Mathews, F.S.; Lidstrom, M.E.
#journal Biochem. Biophys. Res. Commun. (1992) 184:1181-1189
#title The genetic organization of the mau gene cluster of the
facultative autotroph Paracoccus denitrificans.
#cross-references MIM:92272706
#accession JH0660
##molecule_type DNA
##residues 1-417 ##label CHI
##cross-references GB:M90099; NID:g150583; PID:g150585
REFERENCE S51046
#authors van Spanning, R.J.M.; van der Palen, C.J.N.M.; Slotboom,
D.J.; Reijnders, W.N.M.; Stouthamer, A.H.; Duine, J.A.
#journal Eur. J. Biochem. (1994) 226:201-210
#title Expression of the mau genes involved in methylamine
metabolism in Paracoccus denitrificans is under control of
a LysR-type transcriptional activator.
#cross-references MIM:95045590
#accession S51049
##status preliminary
##molecule_type DNA
##residues 1-69 ##label VAN
##cross-references EMBL:U12464; NID:g558801; PID:g558805
GENETICS mauB
#gene
FUNCTION #description catalyzes the oxidation of methylamine to formaldehyde and
ammonia
KEYWORDS oxidoreductase
FEATURE 1-28 #domain (or 1-26) signal sequence #status predicted
#label SIG\
#product (or 27-417) amine dehydrogenase large chain
#status predicted #label AMI
SUMMARY #length 417 #molecular-weight 45440 #checksum 4630

Query Match 33.0%; Score 74; DB 2; Length 417;
Best Local Similarity 40.0%; Pred. No. 8.72e-01;
Matches 8; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 193 VGVVDLEK&FKRMLDVPDC 212
QY 10 IGITELGGHEIKKVLVPGC 28

RESULT 8

```

```

ENTRY      F71663      #type complete
TITLE      sco2 protein precursor (sco2) RP597 - Rickettsia prowazekii
ORGANISM   #formal_name Rickettsia prowazekii
DATE       21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
ACCESSIONS F71663
REFERENCE   A71630
#authors   Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
            Sacheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
            Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
            C.G.
#journal   Nature (1998) 396:133-140
#title     The genome sequence of Rickettsia prowazekii and the origin
            of mitochondria.
#cross-references MUID:99039499
#accession F71663
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-205 #label AND
#cross-references GB:AJ235272; GB:AJ235269; NID:g3861033; PID:el342876;
            PID:g3861132
#experimental_source strain Madrid E
GENETICS   sco2; RP597
#gene      #superfamily immunodominant protein
CLASSIFICATION
SUMMARY    #length 205 #molecular-weight 23644 #checksum 748
Query Match 32.6%; Score 73; DB 2; Length 205;
Best Local Similarity 45.8%; Pred. No. 1.26e+00;
Matches 11; Conservative 6; Mismatches 5; Indels 2; Gaps 2;
Db 126 EYKFNHPKFTSLTG-NEHQIKDV 148
      :||| :|||:| :|||
QY 1 QYIKA-NSKFIGITELGGHEIKV 23
      :||| :|||:| :|||

RESULT     9
ENTRY      B61036      #type complete
TITLE      transforming growth factor beta-5 precursor - African clawed
            frog
ORGANISM   #formal_name Xenopus laevis #common_name African clawed frog
DATE       31-Dec-1993 #sequence_revision 03-Feb-1994 #text_change
ACCESSIONS A34929; B61036
REFERENCE   A34929
#authors   Kondaiah, P.; Sands, M.J.; Smith, J.M.; Fields, A.; Roberts,
            A.B.; Sporn, M.B.; Melton, D.A.
#journal   J. Biol. Chem. (1990) 265:1089-1093
#title     Identification of a novel transforming growth factor-beta
            (TGF-beta5) mRNA in Xenopus laevis.
#cross-references MUID:90110090
#accession A34929
#molecule_type mRNA
#residues  1-382 #label KON
#cross-references GB:J05180; NID:g214821; PIDN:AAA49968.1; PID:g214822
REFERENCE   A61036
#authors   Roberts, A.B.; Rosa, F.; Roche, N.S.; Coligan, J.E.;
            Garfield, M.; Rebert, M.L.; Kondaiah, P.; Danielpour, D.;
            Kehrl, J.H.; Wahl, S.M.; Dawid, I.B.; Sporn, M.B.
#journal   Growth Factors (1990) 2:135-147
#title     Isolation and characterization of TGF-beta2 and TGF-beta5
            from medium conditioned by Xenopus XTC cells.
#cross-references MUID:90253806
#accession B61036
#molecule_type protein
#residues  271-276,'X',278-284,'XX',287-299 #label ROB
CLASSIFICATION
KEYWORDS   #superfamily inhibin
            growth factor
FEATURE    #product transforming growth factor beta-5 #status
            experimental #label MAT
SUMMARY    #length 382 #molecular-weight 44200 #checksum 3471

```

```

Query Match 31.7%; Score 71; DB 2; Length 382;
Best Local Similarity 50.0%; Pred. No. 2.61e+00;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;
Db 100 YAKOVYRPESITELEDHEFK 119
      :| :| :| :| :| :| :| :|
QY 2 YIKANSKFIGITELGGHEIK 21
      :| :| :| :| :| :| :| :|

RESULT     10
ENTRY      A71201      #type complete
TITLE      hypothetical protein PH1878 - Pyrococcus horikoshii
ORGANISM   #formal_name Pyrococcus horikoshii
DATE       14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
            -14-Aug-1998
ACCESSIONS A71201
REFERENCE   A71000
#authors   Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.;
            Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
            Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
            Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
            Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
            A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
            Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
#journal   DNA Res. (1998) 5:55-76
#title     Complete sequence and gene organization of the genome of a
            hyper-thermophilic archaeobacterium, Pyrococcus horikoshii
            OT3.
#cross-references MUID:98344137
#accession A71201
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-393 #label KAW
#cross-references GB:AP000007; NID:g3236134; PID:d1031943; PID:g3258317
#experimental_source strain OT3
#note      this accession replaces an interim accession for a
            sequence replaced by GenBank
GENETICS   PH1878
#gene      #molecular-weight 44906 #checksum 2696
SUMMARY    #length 393 #molecular-weight 44906 #checksum 2696
Query Match 31.3%; Score 70; DB 2; Length 393;
Best Local Similarity 36.4%; Pred. No. 3.74e+00;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
Db 350 QFLIRNAKFLDAIEAENAIK 371
      :| :| :| :| :| :| :| :|
QY 1 QYIKANSKFIGITELGGHEIKK 22
      :| :| :| :| :| :| :| :|

RESULT     11
ENTRY      A64200      #type complete
TITLE      DNA polymerase III beta chain - Mycoplasma genitalium (SGC3)
ORGANISM   #formal_name Mycoplasma genitalium
DATE       17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
            -10-Oct-1997
ACCESSIONS A64200
REFERENCE   A64200
#authors   Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
            R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
            Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
            Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
            Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
            J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
            Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
            C.A.; Venter, J.C.
#journal   Science (1995) 270:397-403
#title     The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession A64200
#status    preliminary; nucleic acid sequence not shown;

```

```

#molecule_type DNA
#residues 1-267 #label TIGR
#cross-references GB:U39679; GB:L43967; NID:g1045668; PID:g1045669;
#experimental_source strain G-37
TIGR:MG001

GENETICS
#gene dnaN
#genetic_code GCG
#start_codon GTG
SUMMARY
#length 267 #molecular_weight 31240 #checksum 4856
Query Match 30.8%; Score 69; DB 2; Length 267;
Best Local Similarity 40.0%; Pred. No. 5.33e+00;
Matches 10; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Db 242 FQGNKYFLITKSEPELQKILVPS 266
QY 3 IRANSKFIGITELGGHEIKKVLVPG 27

RESULT 12
ENTRY F70361 #type complete
TITLE tRNA-pseudouridine synthase (EC 5.4.99.-) truD - Aquifex
ALTERNATE_NAMES aelicus
ORGANISM tRNA pseudouridine 55 synthase
#formal_name Aquifex aelicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
24-Sep-1999
ACCESSIONS F70361
REFERENCE A70300
#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
#journal Nature (1998) 392:353-358
#title The complete genome of the hyperthermophilic bacterium
Aquifex aelicus.
#cross-references MUID:98196666
#accession F70361
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-287 #label AQF
#cross-references GB:AE000703; NID:g2983287; PIDN:AAOC6885.1;
#experimental_source strain VF5

GENETICS
#gene truD
CLASSIFICATION #superfamily Escherichia coli protein P35
KEYWORDS intramolecular transferase; isomerase; tRNA modification
SUMMARY #length 287 #molecular_weight 32259 #checksum 7335
Query Match 30.8%; Score 69; DB 2; Length 287;
Best Local Similarity 59.1%; Pred. No. 5.33e+00;
Matches 13; Conservative 3; Mismatches 4; Indels 2; Gaps 2;

Db 266 DSKFIGIGELKGVLSPKRLIV 287
QY 6 NSKFIGITEL-GGH-EIKKVLV 25

RESULT 13
ENTRY S62836 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain dnaN
ALTERNATE_NAMES - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
ORGANISM hypothetical protein K05_orf380
#formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 24-Aug-1996 #sequence_revision 13-Mar-1997 #text_change
18-Sep-1998
ACCESSIONS S62836; S73479
REFERENCE S62797

#authors Hilbert, H.; Himmelreich, R.; Plagens, H.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:628-639
#title Sequence analysis of 56 kb from the genome of the bacterium
Mycoplasma pneumoniae comprising the dnaA region, the atp
operon and a cluster of ribosomal protein genes.
#cross-references MUID:96177562
#accession S62836
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-380 #label HTL
#cross-references EMBL:U34816; NID:g1209514; PID:g1209517
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkil, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession S73479
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-380 #label HIM
#cross-references EMBL:AE000017; GB:U00089; NID:g1673812; PID:g1673814
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996
GENETICS
#gene dnaN
#genetic_code GCG
KEYWORDS nucleotidyltransferase
SUMMARY #length 380 #molecular_weight 43856 #checksum 8925
Query Match 30.8%; Score 69; DB 2; Length 380;
Best Local Similarity 30.8%; Pred. No. 5.33e+00;
Matches 8; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

Db 354 YFQSNKYFLITSSNNEPELKEILVPS 379
QY 2 YKANSKFIGITELGGHEIKKVLVPG 27

RESULT 14
ENTRY A55188 #type complete
TITLE nucleolar protein NOP2 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein N2428; protein YNL061w; YNAL protein
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change
17-Mar-1999
ACCESSIONS A55188; B55188; S51670; S58717; S62989
REFERENCE A55188
#authors de Beus, E.; Brockenbrough, J.S.; Hong, B.; Aris, J.P.
#journal J. Cell Biol. (1994) 127:1799-1813
#title Yeast NOP2 encodes an essential nucleolar protein with
homology to a human proliferation marker.
#cross-references MUID:95105226
#accession A55188
#molecule_type DNA
#residues 1-618 #label DEA
#cross-references GB:X82656; NID:g576441; PID:g576442
#accession B55188
#molecule_type protein
#residues 207-217:355-369 #label DE2
REFERENCE S51669
#authors Garcia-Barrio, M.T.; Cuesta, R.; Hinnebusch, A.G.; Tamame
Gonzalez, M.
#submission submitted to the EMBL Data Library, December 1994
#accession S51670
#molecule_type DNA
#residues 146-576, 'M', 578-618 #label GAR
#cross-references EMBL:X83512; NID:g603586; PID:g603587

```

```

REFERENCE
#authors      S58711
#journal      Bergez, P.; Doignon, F.; Crouzet, M.
#title        Yeast (1995) 11:967-974
#note         The sequence of a 44 420 bp fragment located on the left arm
               of chromosome XIV from Saccharomyces cerevisiae.
#cross-references MUID:96021608
#accession     S58717
#status        nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues      1-618 #label BER
#cross-references EMBL:U12141; NID:g1314216; PID:g1098483
#note         the nucleotide sequence was submitted to the EMBL Data
               Library, July 1994
REFERENCE      S62975
#authors      Bergez, P.; Doignon, F.; Crouzet, M.
#submission   submitted to the Protein Sequence Database, April 1996
#accession     S62989
#molecule_type DNA
#residues      1-618 #label BEW
#cross-references EMBL:271337; NID:g2253172; PID:e328800; PID:g1301934;
               MIPS:YNL061w
#experimental_source strain S288C
GENETICS
#gene          SGD:NOP2; YNA1
#map_position  14L
#cross-references SGD:S0005005; MIPS:YNL061w
KEYWORDS
#map_position  14L
#length 618 #molecular-weight 69812 #checksum 9325
SUMMARY
Query Match      30.8%; Score 69; DB 2; Length 618;
Best Local Similarity 36.0%; Pred. No. 5.33e+00;
Matches 9; Conservative 5; Mismatches 10; Indels 1; Gaps 1;
Db 326 HYILQAASSFPLVIALDPHENRIL 350
QY 1 QYI-KANSKFIGITELGGHEIKKVL 24
      ||| ||| : ||| : ||| : |||
      QYI-KANSKFIGITELGGHEIKKVL 24

RESULT 15
ENTRY   C64669 #type complete
TITLE   translation elongation factor EF-G - Helicobacter pylori
        (strain 26695)
ORGANISM #formal_name Helicobacter pylori
DATE     09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change
        26-Aug-1999
ACCESSIONS C64669
REFERENCE   A64520
#authors   Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.;
           Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk,
           H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush,
           J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
           Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.;
           McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.;
           Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback, T.R.;
           Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.;
           Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes,
           W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,
           C.M.; Venter, J.C.
#journal   Nature (1997) 388:539-547
#title     The complete genome sequence of the gastric pathogen
           Helicobacter pylori.
#cross-references MUID:97394467
#accession C64669
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
#molecule_type DNA
#residues  1-692 #label TOM
#cross-references GB:AE000625; GB:AE000511; NID:g2314349;
           PIDN:AAD08239.1; PID:g2314354; TIGR:HP1195
CLASSIFICATION #superfamily translation elongation factor G; translation
               elongation factor Tu homology
KEYWORDS      GTP binding; P-loop
FEATURE

```

```

11-138      #domain translation elongation factor Tu homology #label
            ERV
17-24      #region nucleotide-binding motif A (P-loop)\
135-138     #region GTP-binding NKXD motif
SUMMARY     #length 692 #molecular-weight 77021 #checksum 608
Query Match      30.8%; Score 69; DB 2; Length 692;
Best Local Similarity 47.6%; Pred. No. 5.33e+00;
Matches 10; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
Db 229 KYLGGEELSTEEIKKGKAGC 249
QY 8 KFIGITELGGHEIKKVLVPGC 28
      ||| ||| : ||| : ||| : |||
      KFIGITELGGHEIKKVLVPGC 28

Search completed: Sun Apr 2 17:46:51 2000
Job time : 49 secs.

```

MPESRCH

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:53:01 2000; MasPar time 66.11 Seconds
Tabular output not generated. 14.684 Million cell updates/sec

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spiremb12

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 25.207; Variance 28.345; scale 0.889

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	92	86.8	145	5	GROUP 2 ALLERGEN EUR M	4.49e-10
2	63	59.4	157	3	CONSERVED HYPOTHETICAL	2.53e-02
3	61	57.5	292	10	T8011.21 PROTEIN.	7.66e-02
4	57	53.8	77	2	YOJC PROTEIN.	6.56e-01
5	57	53.8	697	2	PUTATIVE TRANSCRIPTION	6.56e-01
6	56	52.8	520	11	SYNTROPHIN-2.	1.11e+00
7	56	52.8	540	4	BETA2-SYNTROPHIN.	1.11e+00
8	56	52.8	635	14	K5 LYASE.	1.11e+00
9	56	52.8	1074	5	T13F2.3 PROTEIN.	1.11e+00
10	56	52.8	1533	4	MELASTATIN 1.	1.11e+00
11	56	52.8	3038	3	LOVASTATIN NONKETIDE	1.11e+00
12	56	52.8	3944	5	C47D12.1 PROTEIN.	1.11e+00
13	55	51.9	250	2	METHIONINE AMINOPEPTID	1.85e+00
14	55	51.9	397	2	NITRATE REDUCTASE (EC	1.85e+00
15	55	51.9	475	5	EG:103B4.2 PROTEIN.	1.85e+00
16	54	50.9	463	10	ALPHA 3 FRUSTULIN (FRA	3.08e+00
17	54	50.9	482	1	DNA POLYMERASE DELTA S	3.08e+00
18	54	50.9	662	11	573K1.1.3 (GAMMA-AMINO	3.08e+00
19	54	50.9	787	14	ORIGIN-BINDING PROTEIN	3.08e+00
20	54	50.9	812	11	GABAB RECEPTOR 1D.	3.08e+00

21	54	50.9	833	11	Q9WV15	573K1.1.4 (GAMMA-AMINO	3.08e+00
22	54	50.9	844	11	Q9WV17	573K1.1.2 (GAMMA-AMINO	3.08e+00
23	54	50.9	844	11	O08621	GABA-BRIB RECEPTOR.	3.08e+00
24	54	50.9	844	4	O96022	GABAB RECEPTOR, SUBUNI	3.08e+00
25	54	50.9	875	11	O920F9	GABAB RECEPTOR 1C.	3.08e+00
26	54	50.9	899	4	O95975	GABAB RECEPTOR, SUBUNI	3.08e+00
27	54	50.9	930	4	O95468	GABABRI PROTEIN.	3.08e+00
28	54	50.9	960	11	O08620	GABA-BRIA RECEPTOR.	3.08e+00
29	54	50.9	960	11	Q9WV18	573K1.1.1 (GAMMA-AMINO	3.08e+00
30	54	50.9	960	11	Q9WU48	GABA-BIA RECEPTOR.	3.08e+00
31	54	50.9	961	4	O95375	GABA-BIA RECEPTOR.	3.08e+00
32	54	50.9	991	11	O920U4	GABAB RECEPTOR SUBTYPE	3.08e+00
33	53	50.0	230	4	O95553	DJ51J12.1.3 (HUMAN ORT	5.09e+00
34	53	50.0	241	5	O95130	HREP1B.	5.09e+00
35	53	50.0	246	4	O95552	DJ51J12.1 (HUMAN ORTHO	5.09e+00
36	53	50.0	325	11	O61110	QUAKING PROTEIN (QKI-7	5.09e+00
37	53	50.0	338	11	O88972	KH DOMAIN RNA BINDING	5.09e+00
38	53	50.0	344	11	O92246	KH DOMAIN RNA BINDING	5.09e+00
39	53	50.0	383	13	O42476	QUAKING PROTEIN HOMOLO	5.09e+00
40	53	50.0	437	10	O23024	TIG11.14 PROTEIN.	5.09e+00
41	53	50.0	438	10	O39495	ALPHA 2 FRUSTULIN.	5.09e+00
42	53	50.0	442	10	O39494	75K MRNA.	5.09e+00
43	53	50.0	632	3	O07845	CHROMOSOME XII READING	5.09e+00
44	52	49.1	110	1	O9YF62	110AA LONG HYPOTHETICA	8.36e+00
45	52	49.1	276	11	Q63020	ALPHA-1-INHIBITOR III	8.36e+00

ALIGNMENTS

RESULT 1
ID O96430 PRELIMINARY; PRT; 145 AA.
AC O96430;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Euroglyphus maynei (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Euroglyphus.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH W., HART B.J., THOMAS W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF047613; AAC82349.1;
DR HSSP: P49278; 1A9V.
SQ SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;

Query Match 86.8%; Score 92; DB 5; Length 145;
Best Local Similarity 85.7%; Pred. No. 4.49e-10;
Matches 12; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 27 HEIKKVMVPCCKGS 40
QY 1 HEIKKVLVPGCHGS 14
|||||:|||||

RESULT 2
ID O74797 PRELIMINARY; PRT; 157 AA.
AC O74797;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN SPBC2D10.03C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;

RA WOOD V., RAJANDREAM M.A., BARRELL B.G., TAYLOR K., HARRIS D.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL031788; CAA21161.1;
 SQ SEQUENCE 157 AA; 17662 MW; FDD493DC CRC32;

Query Match 59.4%; Score 63; DB 3; Length 157;
 Best Local Similarity 61.5%; Pred. No. 2.53e-02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 90 QETKVLVLECRG 102
 : ||||| |:
 QY 1 HEIKKVLVPGCHG 13

RESULT 3
 ID Q92U90 PRELIMINARY; PRT; 292 AA.
 AC Q92U90;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE T8011.21 PROTEIN.
 GN T8011.21

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC eukaryotes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA LIN X., KAUL S., SHEA T.P., FUJII C.Y., SHEN M., VANAKEN S.E.,
 RA BARNSTEAD M.E., MASON T.M., BOWMAN C.L., RONNING C.M., BENITO M.,
 RA CARRERA A.J., CREASY T.H., BUELL C.R., TOWN C.D., NIERMAN W.C.,
 RA FRASER C.M., VENTER J.C.;
 RT "Arabidopsis thaliana chromosome II BAC T8011 genomic sequence."
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC006069; AAD12709.1;
 DR EMBL; AC006069; AAD12709.1;
 SQ SEQUENCE 292 AA; 33005 MW; 9E18A63B CRC32;

Query Match 57.5%; Score 61; DB 10; Length 292;
 Best Local Similarity 63.6%; Pred. No. 7.66e-02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 109 ITRKIVPGCSG 119
 |:::|||||
 QY 3 IKKVLVPGCHG 13

RESULT 4
 ID Q31860 PRELIMINARY; PRT; 77 AA.
 AC Q31860;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE YQJC PROTEIN.
 GN YQJC.

OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RC MEDLINE; 98044033.
 RA KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
 RA AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
 RA BORRIS R., BOURSTIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
 RA BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
 RA CHOI S.K., CODANI J.J., CONNERTON I.F., CONNINGS N.J., DANIEL R.A.,
 RA DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMERSON P.T.,
 RA ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
 RA FRITZ C., FUJITA M., FUJITA Y., FUNA S., GALIZZI A., GALLERON N.,
 RA GHIM S.Y., GLASER P., GORFEAU A., GOLIGHTLY E.J., GRANDI G.,
 RA GUISEPPI G., GUY B.J., HAGA K., HATECH J., HARWOOD C.R., HENAUT A.,

RA HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
 RA JORIS B., KARAWATA D., KASAHARA Y., KLAER-BLANCHARD M., KLEIN C.,
 RA KOBAYASHI Y., KOETER P., KONINGSTEIN G., KROGH S., KUMANO M.,
 RA KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
 RA LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
 RA MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
 RA NOONE D., O'REILLY M., OGAWA K., OGIMAWA A., OUDEGA B., PARK S.H.,
 RA PARRO V., POHL T.M., PORTETELLE D., PORWOLLIK S., PRESCOTT A.M.,
 RA PRESCAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
 RA RIEGER M., RIVOLTA C., ROCHE E., ROCHE M., ROSE M., SADAIE Y.,
 RA SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCORFONE F.,
 RA SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SEROR P., SHIN B.S., SOLDI B.,
 RA SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
 RA TAKEUCHI M., TAKAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
 RA TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
 RA VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
 RA WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
 RA YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA KUNST F., OGASAWARA N., YOSHIKAWA H., DANCHIN A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z99114; CAB13841.1;
 SQ SEQUENCE 77 AA; 8518 MW; B25923A1 CRC32;

Query Match 53.8%; Score 57; DB 2; Length 77;
 Best Local Similarity 66.7%; Pred. No. 6.56e-01;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 14 RTVLVPACH 22
 : |||||:
 QY 4 KKVLVPGCH 12

RESULT 5
 ID Q45419 PRELIMINARY; PRT; 697 AA.
 AC Q45419;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATOR.
 GN MTLR.

OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 7954;
 RC MEDLINE; 96421984.
 RA HENSTRA S.A., TOLNER B., HOEVE DUURKENS R.H., KONINGS W.N.,
 RA ROBILIARD G.T.;
 RT "Cloning, expression, and isolation of the mannitol transport protein
 RT from the thermophilic bacterium Bacillus stearothermophilus."
 RL J. Bacteriol. 178:5586-5591(1996).
 DR EMBL; U18943; AAC44464.1;
 DR PFAM; PF00874; BglG_antitermin; 2.

SQ SEQUENCE 697 AA; 79242 MW; A78A4E30 CRC32;
 Query Match 53.8%; Score 57; DB 2; Length 697;
 Best Local Similarity 58.3%; Pred. No. 6.56e-01;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 549 HSIKEVLAECACH 560
 | || | | | |
 QY 1 HEIKKVLVPGCH 12

RESULT 6 PRELIMINARY; PRT; 520 AA.
 ID Q61235

```

AC Q61235;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE SYNTHROPIN-2.
GN SNTB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RP TISSUE-DIAPHRAGM;
RX MEDLINE; 94000819.
RA ADAMS M.E., BUTLER M.H., DWYER T.M., PETERS M.F., MURNANE A.A.,
RA FROEHNER S.C.;
RT "Two forms of mouse syntrophin, a 58 kd dystrophin-associated protein,
RT differ in primary structure and tissue distribution.";
RL Neuron 11:531-540(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-DIAPHRAGM;
RX MEDLINE; 96029685.
RA ADAMS M.E., DWYER T.M., DOWLER L.L., WHITE R.A., FROEHNER S.C.;
RT "Mouse alpha 1- and beta 2-syntrophin gene structure, chromosome
RT localization, and homology with a discs large domain.";
RL J. Biol. Chem. 270:25859-25865(1995).
DR EMBL; U00678; AAC53060.1; -.
DR HSSP; Q12959; IPDR.
DR MGD; MGI:101771; Sntb2.
DR PFAM; PF00595; PD2; 1.
SQ SEQUENCE 520 AA; 56381 MW; 6FC27580 CRC32;

Query Match 52.8%; Score 56; DB 11; Length 520;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 410 RILVQCHAA 419
QY 5 KVLVPGCHGS 14

RESULT 7
ID Q13425; PRELIMINARY; PRT; 540 AA.
AC Q13425;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE BETA2-SYNTROPHIN.
GN SNT B2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 96162017.
RX AHN A.H., FEENER C.A., GUSSONI E., YOSHIDA M., OZAWA E., KUNKEL L.M.;
RA "The three human syntrophin genes are expressed in diverse tissues,
RT have distinct chromosomal locations, and each bind to dystrophin and
RT its relatives.";
RL J. Biol. Chem. 271:2724-2730(1996).
DR EMBL; U40572; AAC50449.1; -.
DR HSSP; Q12959; IPDR.
DR PFAM; PF00595; PD2; 1.
SQ SEQUENCE 540 AA; 57949 MW; 7016418A CRC32;

Query Match 52.8%; Score 56; DB 4; Length 540;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 430 RILVQCHAA 439
QY 5 KVLVPGCHGS 14

RESULT 8
ID Q09496; PRELIMINARY; PRT; 635 AA.
AC Q09496;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE K5 LYASE.
OS coliphage K5.
OC Viruses.
[1]
RN SEQUENCE FROM N.A.
RP ESCOFFER F., HANFLING P., JANN B., JANN K., ROBERTS I.S.;
RL Submitted (DCL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y10025; CAA71133.1; -.
KW Lyase.
SQ SEQUENCE 635 AA; 66944 MW; 58219280 CRC32;

Query Match 52.8%; Score 56; DB 14; Length 635;
Best Local Similarity 50.0%; Pred. No. 1.11e+00;
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 294 INNVLMSGCDGT 305
QY 3 IKKVLVPGCHGS 14

RESULT 9
ID Q94046; PRELIMINARY; PRT; 1074 AA.
AC Q94046;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE T13F2.3 PROTEIN.
GN T13F2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
RN SEQUENCE FROM N.A.
RP SWINBURNE J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JIER M., JOHNSTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z81122; CAB03354.1; -.
DR PFAM; PF00533; BRCT; 1.
SQ SEQUENCE 1074 AA; 120339 MW; D8BB51C CRC32;

Query Match 52.8%; Score 56; DB 5; Length 1074;
Best Local Similarity 42.9%; Pred. No. 1.11e+00;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 565 HQITHVLVDSRNT 578
QY 1 HEIKKVLVPGCHGS 14

RESULT 10
ID O75560; PRELIMINARY; PRT; 1533 AA.

```

AC 075560; (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE MELASTATIN 1.
 GN MSLN1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-RETINA;
 RA HUNTER J.J., SHAO J., SMUTKO J.S., DUSSAULT B.J., NAGLE D.L.,
 RA WOOLF E.A., HOLMGREN L.M., MOORE K.J., SHYJAN A.W.;
 RT "Chromosomal localization and genomic characterization of the mouse
 RT melastatin gene (Msln1).";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF071787; AAC8000.1;
 SQ SEQUENCE 1533 AA; 174412 MW; E7693DD2 CRC32;
 Query Match 52.8%; Score 56; DB 4; Length 1533;
 Best Local Similarity 46.2%; Pred. No. 1.11e+00;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 Db 1241 DVKTHLVPCEONS 1253
 QY 2 EIRKVLPGCHGS 14
 RESULT 11
 ID Q9Y8A5 PRELIMINARY; PRT; 3038 AA.
 AC Q9Y8A5;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE LOVASTATIN NONKETIDE SYNTHASE.
 GN LOVB.
 OS Aspergillus terreus.
 OC Eukaryota; Fungi; Ascomycota; Euascomycetes; Plecomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 20542;
 RX MEDLINE; 99310964.
 RA HENDRICKSON L., DAVIS C.R., ROACH C., NGUYEN D.K., ALDRICH T.,
 RA MCADA P.C., REEVES C.D.;
 RT "Lovastatin biosynthesis in Aspergillus terreus: characterization of
 RT blocked mutants, enzyme activities and a multifunctional polyketide
 RT synthase gene.";
 RL Chem. Biol. 6:429-439(1999).
 DR EMBL: AF151722; AAD39830.1;
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 KW Transferase.
 SQ SEQUENCE 3038 AA; 335001 MW; 90A5331A CRC32;
 Query Match 52.8%; Score 56; DB 3; Length 3038;
 Best Local Similarity 54.5%; Pred. No. 1.11e+00;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 Db 2046 ORMKVLMPCG 2056
 QY 1 HEIRKVLVPGC 11
 RESULT 12
 ID Q18667 PRELIMINARY; PRT; 3944 AA.
 AC Q18667;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-JAN-1999 (Tremblrel. 09, Last annotation update)
 DE C47D12.1 PROTEIN.
 GN C47D12.1.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA GAJADSTY S.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z69902; CAA93765.1;
 SQ SEQUENCE 3944 AA; 452752 MW; C44B1EB5 CRC32;
 Query Match 52.8%; Score 56; DB 5; Length 3944;
 Best Local Similarity 50.0%; Pred. No. 1.11e+00;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Db 1991 VKKILIECH 2000
 QY 3 IKKVLVPGCH 12
 RESULT 13
 ID Q9X117 PRELIMINARY; PRT; 250 AA.
 AC Q9X117;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE METHIONINE AMINOPEPTIDASE (EC 3.4.11.18) (MAP).
 GN TM1478.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 99287316.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
 RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RT "Evidence for lateral gene transfer between Archaea and bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 395:323-329(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
 RA HAFT D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
 RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
 RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
 RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
 RA SMITH H.O., VENTER J.C., FRASER C.M.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REMOVES THE AMINO-TERMINAL METHIONINE FROM NASCENT
 CC PROTEINS...
 CC -1- CATALYTIC ACTIVITY: L-METHIONYLPEPTIDE + H(2)O = L-METHIONINE +
 CC PEPTIDE.
 CC -1- COFACTOR: COBALT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M24A; ALSO KNOWN AS THE
 CC MAP FAMILY.
 DR EMBL; AE001798; AAD36544.1; -.

DR PROSITE; PS00680; MAP.1; 1.
 KW Aminopeptidase; Hydrolase; Cobalt.
 SQ SEQUENCE 250 AA; 27489 MW; BA9A64CF CRC32;

Query Match 51.9%; Score 55; DB 2; Length 250;
 Best Local Similarity 60.0%; Pred. No. 1.85e+00;
 Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 25 REVRKVIIVPG 34
 QY 1 HEIKKVLVPG 10

RESULT 14
 ID O33732 PRELIMINARY; PRT; 397 AA.
 AC O33732;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JAN-1998 (TREMBLrel. 05, Last annotation update)
 DE NITRATE REDUCTASE (EC 1.7.99.4) (FRAGMENT).
 OS Shewanella putrefaciens.
 OC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 OC Shewanella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCIMB400;
 RA GORDON E.H.J., PIKE A.D., FISCHER H., CHAPMAN S.K., REID G.A.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ000006; CAA03851.1;
 KW Oxidoreductase.
 FT NON_TER 397
 SQ SEQUENCE 397 AA; 42187 MW; 5F0AF4AC CRC32;

Query Match 51.9%; Score 55; DB 2; Length 397;
 Best Local Similarity 60.0%; Pred. No. 1.85e+00;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 290 KVLVATSCQG 299
 QY 4 KVLVPGCHG 13

RESULT 15
 ID O46038 PRELIMINARY; PRT; 475 AA.
 AC O46038;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE EG:103B4.2 PROTEIN.
 GN EG:103B4.2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MURPHY L., HARRIS D., BARRELL B.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BENOS P.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL009193; CAA15692.1;
 DR FLYBASE; FBgn0023550; EG:103B4.2.
 SQ SEQUENCE 475 AA; 53756 MW; FBFF0FEE CRC32;

Query Match 51.9%; Score 55; DB 5; Length 475;
 Best Local Similarity 55.6%; Pred. No. 1.85e+00;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 343 KNLMPGCH 351
 QY 4 KVLVPGCH 12

Search completed: Sun Apr 2 17:54:11 2000
 Job time : 70 secs.

MPERCH_PP

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:58:10 2000; MasPar time 75.60 Seconds
125.641 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-3

Description: (1-137) from US09362731.pep

Perfect Score: 1057

Sequence: 1 DQYKANSKFIGITELGGQY.....FGCHGSEPCIIHRGKPF 137

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprenb112

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp_mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp_unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 34.307; Variance 56.239; scale 0.610

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	10.0	145	5	GROUP 2 ALLERGEN EUR M	1.65e-05
2	94	8.9	1792	13	LAMININ BETA 2-LIKE CH	2.41e-03
3	90	8.5	1000	2	CATION EFFLUX (ACRB/AC	1.19e-02
4	89	8.4	3871	5	ZC116.3 PROTEIN.	1.76e-02
5	88	8.3	313	5	EXTRACELLULAR MATRIX P	2.60e-02
6	88	8.3	379	5	GLUE PROTEIN.	2.60e-02
7	87	8.2	988	6	GASTRIC MUCIN (FRAGMEN	3.83e-02
8	83	7.9	373	5	FMRFAMIDE (FRAGMENT).	1.77e-01
9	83	7.9	595	10	NEOXANTHIN CLEAVAGE EN	1.77e-01
10	84	7.9	3623	11	INTRINSIC FACTOR-B12 R	1.21e-01
11	82	7.8	127	5	FMRFAMIDE (FRAGMENT).	2.58e-01
12	82	7.8	925	14	POSS. PRECURSOR POLYPE	2.58e-01
13	82	7.8	929	14	POSS. PRECURSOR POLYPE	2.58e-01
14	82	7.8	1193	13	C-SERATE-1 PROTEIN (FR	2.58e-01
15	81	7.7	177	5	ORF-RF1.	3.74e-01
16	81	7.7	3623	4	INTRINSIC FACTOR-B12 R	3.74e-01
17	80	7.6	356	6	UNKNOWN MRNA, PARTIAL	5.41e-01
18	79	7.5	1081	4	MUC5AC PROTEIN (FRAGME	7.82e-01
19	78	7.4	363	14	THYMIDINE KINASE.	1.13e+00
20	78	7.4	752	13	NOTCH RECEPTOR PROTEIN	1.13e+00

21	78	7.4	1042	4	Q13792	1.13e+00
22	78	7.4	1219	11	Q63722	1.13e+00
23	78	7.4	2704	5	Q97458	1.13e+00
24	77	7.3	199	2	Q92E53	1.62e+00
25	77	7.3	398	2	Q9X7N6	1.62e+00
26	77	7.3	750	4	Q08424	1.62e+00
27	77	7.3	1247	2	Q07910	1.62e+00
28	77	7.3	1764	5	Q02650	1.62e+00
29	77	7.3	4289	4	Q78530	1.62e+00
30	76	7.2	107	10	Q24233	1.62e+00
31	76	7.2	107	10	Q24232	2.31e+00
32	76	7.2	154	10	Q39814	2.31e+00
33	76	7.2	187	2	Q9X5H4	2.31e+00
34	76	7.2	421	3	Q13960	2.31e+00
35	76	7.2	443	2	Q92422	2.31e+00
36	76	7.2	599	4	Q9Y2N9	2.31e+00
37	76	7.2	894	11	Q88715	2.31e+00
38	75	7.1	128	5	Q61610	3.30e+00
39	75	7.1	263	4	Q99740	3.30e+00
40	75	7.1	280	14	Q89642	3.30e+00
41	75	7.1	1218	4	Q15816	3.30e+00
42	75	7.1	1218	4	Q14902	3.30e+00
43	75	7.1	1218	4	Q15122	3.30e+00
44	75	7.1	1227	4	Q78504	3.30e+00
45	75	7.1	2026	4	Q00468	3.30e+00

ALIGNMENTS

RESULT 1
ID Q96430 PRELIMINARY; PRT; 145 AA.
AC Q96430;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Eukaryotes; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Eukaryophus.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH W., HART B.J., THOMAS W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047613; AAC82349.1;
DR HSSP; P49278; IA9V.
SQ SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;

Query Match 10.0%; Score 106; DB 5; Length 145;
Best Local Similarity 75.0%; Pred. No. 1.65e-05;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 36 GCGSEPCVHRGTAF 51
QY 52 GCHGSEPCIIHRGKPF 67
|| |||||:||||:|

RESULT 2
ID Q57484 PRELIMINARY; PRT; 1792 AA.
AC Q57484;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE LAMININ BETA 2-LIKE CHAIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93015947.
RA O'REAR J.J.;
RT "A novel laminin B1 chain variant in avian eye."

01-NOV-1999	(TrEMBLrel. 12, Last annotation update)
ZC116.3	PROTEIN.
CENKRN	
Eukaryotic	elegans.
Karyota; Metazoa;	Nematoda; Secernentea; Rhabditia; Rhabditiida;
Rhabditina;	Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]	
SEQUENCE FROM N.A.	
BURTON J.;	
Submitted (JUN-1996)	to the EMBL/GenBank/DDBJ databases.
[2]	
SEQUENCE FROM N.A.	
MEDLINE; 94130718.	
WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,	
BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,	
CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,	
GARDNER A., GREEN P., HAWKINS J., HILLIER L., JISER M., JOHNSTON L.,	
JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,	
LIGHTNING J., LOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,	
PARKSONS J., PERCY C., RIFFEN L., ROOPRA A., SAUNDERS D., SHOWNEEN R.,	
SAMDALDN N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,	
THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,	
WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;	
*2.2 kb of contiguous nucleotide sequence from chromosome III of C.	
elegans.*;	
Nature 368:32-38(1994).	
[3]	
SEQUENCE FROM N.A.	
SMYE R.;	
Submitted (JUN-1996)	to the EMBL/GenBank/DDBJ databases.
EMBL; 274473; CA98952.1;	-
EMBL; 274046; CA98952.1;	JOINED.
EMBL; 274046; CA98957.1;	-
EMBL; 274473; CA98957.1;	JOINED.
HSP; P00740.1; XA.	
PROSITE; PS01187; EGF_CA; 3.	
PFAM; PF00431; COB; 13.	
PFAM; PF00008; EGF; 7.	
Glycoprotein; EGF-like domain.	
SEQUENCE 3871 AA; 433659 MW; 0B3CAF62 CRC32;	
Query Match 8.4%; Score 89; DB 5; Length 3871;	
Best Local Similarity 22.6% ; Pred.No.1.76e-02;	
Matches 21; Conservative 20; Mismatches 47; Indels 5; Gaps	
Ddb	56 GLVKWKTRQITGIINKLIAINACDPKNCSGGTCTPSFGAKTCICLPHTGTTCGAIDIDE 115
Qy	18 GOYIKANSKFGITELSSCHSGSEPCLIIHRGPFGCGHSE-PCIIHKRKPFSCHGS-EP 75
Ddb	116 CSVYNGTT-AGCONQTGINNRG-GFECQCQSQ 146
Qy	76 CLIIHRKGPGGCGHGSEPCIIHRGKPFSS-CHGS 107
RESULT 5	
ID Q24330 PRELIMINARY; PRU; 313 AA.	
Q24330;	
DT Q1-NOV-1996 (TrEMBLrel. 01, Created)	
DT Q1-NOV-1996 (TrEMBLrel. 01, Last sequence update)	
DE Q1-NOV-1998 (TrEMBLrel. 08, Last annotation update)	
DE EXTRACELLULAR MATRIX PROTEIN B (FRAGMENT).	
GN ECMB.	
OOS Dictyostelium minutum.	
OC Eukaryota; Dictyosteliida; Dictyostelium.	
[1]	
RN SEQUENCE FROM N.A.	
RP STRAIN=71-2;	
RX MEDLINE; 94336717.	
RA VAN ES S., NIEUWHUIJSSEN B.W., LENOUVEL F., VAN DEUSEN E.M.,	
SCHAAP P.;	
"Universal signals control slime mold stalk formation."	
RT Proc. Natl. Acad. Sci. U.S.A. 91:8219-8223(1994).	
DR EMBL; X78948; CAA55545.1; -	

```

DR EMBL; Z29565; CAA82671.1; -.
DR EMBL; X76203; CAA53797.1; -.
DR FLYBASE; FBgn0C11269; Dvir\Lgp3.
FT DOMAIN 36 39 POLY-THR.
SQ SEQUENCE 379 AA; 41083 MW; D832E6B3 CRC32;

Query Match      8.3%; Score 88; DB 5; Length 379;
Best Local Similarity 24.0%; Pred. No. 2.60e-02;
Matches 23; Conservative 22; Mismatches 46; Indels 5; Gaps 5;

Db    80 TKPTTTRTKP-TTTRTKKPTTTRRKPK-TTTRTKPTTTRRTPKP-TTTRTKPTTTRRT 136
QY    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
       39 SEPCHHGGPFGGCGHGSEPCIIHRGKFSSGSGSEPCIIHRGKPGFGCHGSGSEPCIHRG 98
       || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    137 KP-TTTRTKPTTTRTKP-TTTRTKTKPTTTRTKPK 170
QY    || : : : : : : || : : : : : : : : : : : : : : : : :
       99 KPFSSCHGSEPCIIHRGKPGFGCHGSGSEPCIIHRGKP 134

RESULT 7
ID ID O97867 PRELIMINARY; PRT; 988 AA.
AC O97867;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE GASTRIC MUCIN (FRAGMENT).
GN MUC5AC.
OS Sus scrofa domestica (domestic pig).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
[1]
RN RP SEQUENCE OF 35-109 FROM N.A.
RC TISSUE=GASTRIC EPITHELIUM;
RX MEDLINE: 95275264.
RA TURNER B.-S., BHASKAR K.R., HADZOPOULOU-CIADARAS M., SPECIAN R.D.,
LA MONT J.T.;
RT "Isolation and characterization of cDNA clones encoding pig gastric
   mucin.";
RL Biochem. J. 308:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=GASTRIC EPITHELIUM;
RA TURNER B.-S., BHASKAR K.R., HADZOPOULOU-CIADARAS M., LA MONT J.T.;
RT "Cysteine-rich regions of pig gastric mucin contain cysteine knot and
   von Willebrand factor domains at the 3'-carboxyl terminal: similarity
   to human and rat MUC5AC.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF054584; AAD19833.1; -.
FT NON_TER 1.
SQ SEQUENCE 988 AA; 105264 MW; 2D3E3F11 CRC32;

Query Match      8.2%; Score 87; DB 6; Length 988;
Best Local Similarity 22.2%; Pred. No. 3.83e-02;
Matches 22; Conservative 25; Mismatches 47; Indels 5; Gaps 4;

Db    452 VTLPGOPPCAPPRLTVVETPTPTSCPPSPICQLILSEVPAPCAETPPWPFQGCVF 511
QY    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
       28 IGTEUSSCHGSEP--CIHRGKPGFGCHGSGSEPCIIHRGKPFSSCHGSEPC-IIHRGRP 84
       || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db    512 DHCHMPTDVLCSGLELYAALCASLGVCIDWRGRNHTC 550
QY    || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
       85 GGCHGSEPCIIHRG-KPFSS-CHGSEPCIIHRGKPGGC 121

RESULT 8
ID ID Q16964 PRELIMINARY; PRT; 373 AA.
AC Q16964;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DI 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE FMRFAMIDE (FRAGMENT).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anasoidae;
```

<hr/>						
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 98148073.					
RA	MOESTRUP S.K., KOZYRAKI R., KRISTIANSEN M., KAYSEN J.H.,					
RA	RASMUSSEN H.H., BRAULT D., PONTILLON F., GODA F.O., CHRISTENSEN E.I.,					
RA	HAMMOND T.G., VERKROUST P.J.;					
RT	"The intrinsic factor-vitamin B12 receptor and target of teratogenic					
RT	antibodies is a metal-binding peripheral membrane protein with					
RT	homology to developmental proteins.";					
RL	J. Biol. Chem. 273:5235-5242(1998).					
DR	EMBL; AF022247; AAC71661.1; -.					
DR	HSSP; P00740; IIXA.					
DR	PROSITE; PS00010; ASX_HYDROXYL; 3.					
DR	PROSITE; PS01187; EGF CA; 4.					
DR	PFAM; PF00431; CUB; 27.					
DR	PFAM; PF00008; EGF; 7.					
KW	Signal; Glycoprotein; EGF-like domain.					
FT	SIGNAL 1 20 POTENTIAL.					
FT	CHAIN 21 3623 INTRINSIC FACTOR-B12 RECEPTOR.					
SQ	SEQUENCE 3623 AA; 398981 MW; ADF804DC CRC32;					
Query Match	7.9%; Score 84; DB 11; Length 3623;					
Best Local Similarity	25.3%; Pred.No. 1.2le-01;					
Matches	14; Conservative 13; Mismatches 22; Indels 6; Gaps 5;					
Db	171 CVVYSGTFPG-COSGSTCVNTVG-SPR-CDCTPDVTGYPOCASKYNDEOGSKQLC 222					
	::					
QY	42 CIITHRGKPFGCGHSGSEPCIHRGKPFSSCH-GSEPCIIHRGKFPFGCC-HGSEP-C 93					
	::					
RESULT	11					
ID	Q16963 PRELIMINARY; PRT; 127 AA.					
AC	Q16963;					
DT	01-NOV-1996 (TREMBLrel. 01, Created)					
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)					
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)					
DE	FMRFAMIDE (FRAGMENT).					
OS	Aplysia californica (California sea hare).					
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspiidea;					
CC	Aplysidae; Aplusia.					
CN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=ABDOMINAL GANGLION;					
RX	MEDLINE; 87132918.					
RA	TAUSSIG R., SCHELLER R.H.;					
RT	"The Aplysia fmramide gene encodes sequences related to mammalian					
RT	brain peptides.";					
RL	DNA 5:453-461(1986).					
DR	EMBL; L29473; AAB59170.1; -.					
DR	PFAM; PF01581; FARP; 7.					
KW	Neuropeptide.					
FT	NON_TER 1 1					
FT	CHAIN 4 7 POTENTIAL.					
FT	CHAIN 20 23 POTENTIAL.					
FT	CHAIN 36 39 POTENTIAL.					
FT	CHAIN 52 55 POTENTIAL.					
FT	CHAIN 68 71 POTENTIAL.					
FT	CHAIN 84 87 POTENTIAL.					
FT	CHAIN 91 94 POTENTIAL.					
SQ	SEQUENCE 127 AA; 14697 MW; CFE78DE0 CRC32;					
Query Match	7.8%; Score 82; DB 5; Length 127;					
Best Local Similarity	25.0%; Pred.No. 2.58e-01;					
Matches	23; Conservative 22; Mismatches 42; Indels 5; Gaps 5;					
Db	3 RFMRFGKSVDGS-VDKRFMRFGKSVGTDD-VDKRFMRFGKSLGTTDD-VDKRFMRFGKSL 59					
	::					
QY	46 RGXPFGCGHGSEPCIHRGKPFSSCHGSEPCIHRGKPTGGCHGSEPCIHRGKPFSSCH 105					
	::					
Db	60 GTED-VDKRFMRFGKSLGTED-VDKRFMRFGK 89					
	::					
QY	106 GSEPCIHRGKPFGGCHGSEPCIHRGKPFGR 137					
	::					

```

RC TISSUE=SPINAL CORD;
RX MEDLINE; 96175595.
RA MYAT A., HENRIQUE D., ISH-HOROWICZ D., LEWIS J.;
RT "A chick homologue of Serrate and its relationship with Notch and
RT Delta homologues during central neurogenesis.";
RL Dev. Biol. 174:233-247(1996).
DR EMBL; X95283; CAA64604.1; -.
DR HSSP; P00740; IIXA.
DR PROSITE; PS00010; ASX_HYDROXYL; 10.
DR PROSITE; PS01167; EGF_CA. 8.
DR PFAM; PF01414; DSL; 1.
DR PFAM; PF00008; EGF; 14.
DR KW Glycoprotein; EGF-like domain.
FT NON_TER 1
SQ SEQUENCE 1193 AA; 131039 MW; 55E5FCD1 CRC32;

Query Match 7.8%; Score 82; DB 13; Length 1193;
Best Local Similarity 30.2%; Pred. No. 2.58e-01;
Matches 13; Conservative 10; Mismatches 16; Indels 4; Gaps 4;

Db 871 C-GPRPC11H-AGHNECPAGHACVPKEDHCFTHPCAAGEC 911
| : ||||| : | : : : | : : : | : | : | : | : |
QY 36 CHGSEPC11HRGKPFPGCGHGSEPCI-IHRGKPFSS-SCHGSEPC 76

RESULT 15
ID Q26577 PRELIMINARY; PRT; 177 AA.
AC Q26577;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ORF-RF1.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86287295.
RA BOBEK L., REKOSH D.M., VAN KEULEN H., LOVERDE P.T.;
RT "Characterization of a female-specific cDNA derived from a
RT developmentally regulated mRNA in the human blood fluke Schistosoma
RT mansoni.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:5544-5548(1986).
DR EMBL; M14309; AAA74695.1; -.
DR HSSP; P30129; 4DPV.
SQ SEQUENCE 177 AA; 16332 MW; F17CF2EF CRC32;

Query Match 7.7%; Score 81; DB 5; Length 177;
Best Local Similarity 32.1%; Pred. No. 3.74e-01;
Matches 26; Conservative 18; Mismatches 26; Indels 11; Gaps 11;

Db 31 GGCYGS-D-CDSGYDSGYGGCGTGS-D-CGGYGGYGGGCGSGD-CGNY-GGGYGGDCNG 86
||| | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 51 GCGHSEPC11HRGK-P-FSS-CHGSEPC11HRGKPFGG-CHGSEPC11HRGKPFSS-CHG 106
: : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 87 GD-CGNY-GGGYGGGNGGG-C 104
: : | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
QY 107 SEPC11HRGKPFPGCGHGSEPC 127

Search completed: Sun Apr 2 17:59:30 2000
Job time : 80 secs.

```

MPSEARCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:57:16 2000; MasPar time 30.92 Seconds
Tabular output not generated. 132.305 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.p
Perfect Score: 1057
Sequence: 1 DOYKANSKFIGITELGGQY.....FGGCHGSEPCIIHRGKPFPSR 137

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot

Statistics: Mean 35.712; Variance 57.531; scale 0.621

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	133	12.6	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	6.73e-11
2	132	12.5	146	1	DER2_DERPT MITE ALLERGEN DER P 2	1.07e-10
3	104	9.8	1314	1	TETX_CLOTE TETANUS TOXIN PRECURSO	2.67e-05
4	91	8.6	647	1	ZG48_XENLA GASTRULA ZINC FINGER P	5.41e-03
5	89	8.4	1798	1	LMB2_HUMAN LAMININ BETA-2 CHAIN P	1.18e-02
6	88	8.3	713	1	DDX4_RAT DEAD BOX PROTEIN 4 (VA	1.74e-02
7	86	8.1	552	1	FMR2_APLCA FMRFAMIDE NEUROPEPTIDE	3.76e-02
8	85	8.0	435	1	2022_XENLA OOCYTE ZINC FINGER PRO	5.50e-02
9	85	8.0	1046	1	PSTA_DICD1 PRESTALK PROTEIN PRECU	5.50e-02
10	85	8.0	1801	1	LMB2_RAT LAMININ BETA-2 CHAIN P	5.50e-02
11	84	7.9	309	1	FMR1_APLCA FMRFAMIDE NEUROPEPTIDE	8.02e-02
12	82	7.8	1046	1	VGNM_CPMV GENOME POLYPROTEIN M	1.69e-01
13	79	7.5	2703	1	NOTC_DROME NEUROGENIC LOCUS NOTCH	5.08e-01
14	78	7.4	94	1	SAPE_SARPE SAPECN PRECURSOR	7.27e-01
15	78	7.4	94	1	DEF1_PROTE PHORMICIN PRECURSOR (I	7.27e-01
16	78	7.4	177	1	EGG3_SCHMA EGGSHELL PROTEIN PRECU	7.27e-01
17	78	7.4	177	1	EGG2_SCHMA EGGSHELL PROTEIN PRECU	7.27e-01
18	78	7.4	364	1	KITH_ILTVT THYMIDINE KINASE (EC 2	7.27e-01
19	78	7.4	637	1	DDX4_MOUSE DEAD BOX PROTEIN 4 (VA	7.27e-01
20	78	7.4	732	1	ACPH_HUMAN ACYLAMINO-ACID-RELEASE	1.04e-00
21	77	7.3	261	1	YH86_SYN3 HYPOTHETICAL 29.3 KD P	1.04e-00
22	77	7.3	757	1	COMP_HUMAN CARTILAGE OLIGOMERIC M	1.04e-00
23	76	7.2	755	1	COMP_RAT CARTILAGE OLIGOMERIC M	1.48e+00

24	76	7.2	1799	1	LMB2_MOUSE LAMININ BETA-2 CHAIN P	1.48e+00
25	75	7.1	141	1	ALG2_TYRPU GROUP 2 ALLERGEN PRECU	2.10e+00
26	74	7.0	1147	1	CGAL_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
27	74	7.0	1182	1	CGA2_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
28	74	7.0	1186	1	CGA0_HELPY CYTOTOXICITY ASSOCIATE	2.97e+00
29	73	6.9	362	1	AROG_HAEIN PHOSPHO-2-DEHYDRO-3-DE	4.19e+00
30	73	6.9	1056	1	MUC5_HUMAN TRACHEOBRONCHIAL MUCIN	4.19e+00
31	73	6.9	1064	1	SYIC_SCHPO ISOLEUCYL-TRNA SYNTHET	4.19e+00
32	73	6.9	1713	1	LMA3_HUMAN LAMININ ALPHA-3 CHAIN	4.19e+00
33	73	6.9	4660	1	LRP2_RAT LOW-DENSITY LIPOPROTEI	4.19e+00
34	72	6.8	141	1	LEP1_LEPDS MITE ALLERGEN LEP D 1	5.88e+00
35	72	6.8	216	1	RL3_RICPR 50S RIBOSOMAL PROTEIN	5.88e+00
36	72	6.8	245	1	MTA_METJA TETRAHYDROMETHANOPTERI	5.88e+00
37	72	6.8	305	1	YHDW_ECOLI HYPOHETICAL AMINO-ACI	5.88e+00
38	72	6.8	616	1	ACOC_SOLTU ACONITATE HYDRATASE, C	5.88e+00
39	72	6.8	732	1	ACPH_RAT ACYLAMINO-ACID-RELEASE	5.88e+00
40	72	6.8	1630	1	MSPI_PLAFK MEROZOITE SURFACE PROT	5.88e+00
41	72	6.8	1639	1	MSPI_PLAFM MEROZOITE SURFACE PROT	5.88e+00
42	72	6.8	1964	1	NTC4_MOUSE NEUROGENIC LOCUS NOTCH	5.88e+00
43	72	6.8	3075	1	LMA1_HUMAN LAMININ ALPHA-1 CHAIN	5.88e+00
44	71	6.7	196	1	FMR4_APLCA FMRFAMIDE NEUROPEPTIDE	8.22e+00
45	71	6.7	488	1	ZF92_MOUSE ZINC FINGER PROTEIN 92	8.22e+00

ALIGNMENTS

RESULT 1
ID DEF2_DERFA STANDARD; PRT; 146 AA.
AC Q00835; P39672; Q26359;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
GN DERF2.
OS Dermatophagoides farinae (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 91291341.
RA YUUKI T., OKUMURA Y., ANDO T., YAMAKAWA H., SUKO M., HAIDA M.,
RA OKUDAIRA H.;
RT Cloning and expression of cDNA coding for the major house dust mite
RT allergen Der f II in Escherichia coli.;
RL Agric. Biol. Chem. 55:1233-1238(1991).
RN [2]
RP SEQUENCE OF 4-146 FROM N.A.
RX MEDLINE; 94256850.
RA OKUDAIRA H.;
RT "Molecular biology of mite antigens.";
RA Aterugi 43:435-440(1994).
RN [3]
RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
RX MEDLINE; 93283958.
RA NISHIYAMA C., YUUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
RT "Determination of three disulfide bonds in a major house dust mite
RT allergen, Der f II";
RL Int. Arch. Allergy Immunol. 101:159-166(1993).
RN [4]
RP PARTIAL SEQUENCE OF 18-52.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATTS-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [5]
RP STRUCTURE BY NMR.
RX MEDLINE; 98079068.
RA ICHIKAWA S., HATANAKA H., YUUKI T., IWAMOTO N., KOJIMA S.,
RA NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
RT "Solution structure of Der f 2, the major mite allergen for atopic

RT diseases.";
 RL J. Biol. Chem. 273:356-360(1998).
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D10447; BAA01239.1; -;
 DR EMBL; D10448; BAA01240.1; -;
 DR EMBL; D10449; BAA01241.1; -;
 DR EMBL; S70378; AAB30829.1; -;
 DR PIR; PS0417; PS0417.
 DR PDB; 1AHK; 08-APR-98.
 DR PDB; 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT VARIANT 90 95
 FT VARIANT 105 105 M -> V (IN CLONE 1).
 FT VARIANT 105 105 I -> A (IN CLONE 11).
 FT VARIANT 128 128 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
 SQ SEQUENCE 146 AA; 15802 MW; 72623F23 CRC32;
 Query Match 12.6%; Score 133; DB 1; Length 146;
 Best Local Similarity 88.2%; Pred. No. 6.73e-11;
 Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Db 37 GCHGSEPCIIHRGKPF 53
 QY 52 GCHGSEPCIIHRGKPF 68
 RESULT 2
 ID DER2_DERPT STANDARD; PRT; 146 AA.
 AC P49278;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
 GN DERP2.
 OS Dermatophagoides pteronyssinus (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
 OC Dermatophagoides.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 90256301.
 RA CHUA K.Y., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
 RA THOMAS W.R.;
 RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
 RT plaque immunoassay.";
 RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
 RN [2]
 RP PARTIAL SEQUENCE OF 18-57.
 RX MEDLINE; 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATTS-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [3]

RP STRUCTURE BY NMR.
 RX MEDLINE; 98409423.
 RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
 RT "Tertiary structure of the major house dust mite allergen Der p 2:
 RT sequential and structural homologies.";
 RL Biochemistry 37:12707-12714(1998).
 CC -1- SIMILARITY: TO MITE ALLERGEN LEP D I.
 DR PDB; 1A9V; 14-OCT-98.
 KW Allergen; Signal; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER P 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;
 Query Match 12.5%; Score 132; DB 1; Length 146;
 Best Local Similarity 100.0%; Pred. No. 1.07e-10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 37 GCHGSEPCIIHRGKPF 52
 QY 52 GCHGSEPCIIHRGKPF 67
 RESULT 3
 ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: Primary structure, expression in E. coli, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911;
 RX MEDLINE; 87040747.
 RA FAIRWEATHER N.F., LYNESS V.A.;
 RT "The complete nucleotide sequence of tetanus toxin.";
 RL Nucleic Acids Res. 14:7809-7812(1986).
 RN [3]
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE; 86085672.
 RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in Escherichia coli.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE; 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE; 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]

RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE: 93010948.
RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
RA MONTECUCCO C.;
RT "Tetanus toxin is a zinc protein and its inhibition of
RT neurotransmitter release and protease activity depend on zinc.";
RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE: 93063293.
RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
RA DASGUPTA B.R., MONTECUCCO C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE: 97475217.
RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
RA SAX M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin.";
RL Nat. Struct. Biol. 4:788-792(1997).
CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC BOND OF SYNAPTOSOMAL-2.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
CC SYNAPTOSOMAL-2.
CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GANGLIOSIDE RECEPTORS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEINASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04336; CAA28033.1; -
DR EMBL: M12739; AAA23282.1; -
DR EMBL: X06214; CAA29564.1; -
DR PIR: A25689; BTCLTN.
DR PDB: 1AF9; 29-APR-98.
DR PDB: 1A8D; 14-OCT-98.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
FT 3D-structure. 0 0
FT INIT_MET 0 0
FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 233 233 BY SIMILARITY.
FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT DISULFID 438 466 INTERCHAIN.
FT DISULFID 1076 1092
SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;

Query Match 9.88; Score 104; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 2.67e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYKANSKFCITEL 843
QY 2 QYKANSKFCITEL 16

RESULT 4
ID ZG48_XENLA STANDARD; PRT; 647 AA.
AC P18723;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE GASTRULA ZINC FINGER PROTEIN XLCGF48.2 (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE OF 1-339 FROM N.A.
RX MEDLINE: 89345612.
RA KNOCHHEL W., POTING A., KOSTER M., EL BARADI T., NIETZELD W.,
RA BOUMMEESTER T., PIELER T.;
RT "Evolutionary conserved modules associated with zinc fingers in
RT Xenopus laevis.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6097-6100(1989).
RN [2]
RP SEQUENCE OF 312-647 FROM N.A.
RX MEDLINE: 90040698.
RA NIETZELD W., EL-BARADI T., MENTZEL H., PIELER T., KOESTER M.,
RA POTING A., KNOCHHEL W.;
RT "Second-order repeats in Xenopus laevis finger proteins.";
RL J. Mol. Biol. 208:639-659(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M25871; AAA50018.1; -
DR PIR: F33282; F33282.
DR PIR: S06573; S06573.
DR HSP: P08047; ISP2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2; 12.
DR PFAM: PF00096; zf-C2H2; 12.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 317 339 C2H2-TYPE.
FT ZN_FING 345 367 C2H2-TYPE.
FT ZN_FING 373 395 C2H2-TYPE.
FT ZN_FING 401 423 C2H2-TYPE.
FT ZN_FING 429 451 C2H2-TYPE.
FT ZN_FING 457 479 C2H2-TYPE.
FT ZN_FING 485 507 C2H2-TYPE.
FT ZN_FING 513 535 C2H2-TYPE.
FT ZN_FING 540 563 C2H2-TYPE.
FT ZN_FING 541 563 C2H2-TYPE.
FT ZN_FING 569 591 C2H2-TYPE.
FT ZN_FING 625 647 C2H2-TYPE.
FT NON_TER 647 647
SQ SEQUENCE 647 AA; 73054 MW; C969B307 CRC32;

Query Match 8.68; Score 91; DB 1; Length 647;
Best Local Similarity 36.7%; Pred. No. 5.41e-03;
Matches 22; Conservative 9; Mismatches 24; Indels 5; Gaps 5;

Db 412 KFSRLTEHRCHTGEFFVCSECKSPAQSPQLDHRHRTGEPFCS-CSDGKCFTORA 470
QY 26 KFTGITELSSCH-GSEPCI-IHRGKPGGCHGSEPCI-IHRG-KPFSSCHGSEPCIHRG 81

01-AUG-1988 (Rel. 08, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-JUL-1999 (Rel. 38, Last annotation update)
EMRFAMIDE NEUROPEPTIDE PRECURSOR (CLONE EMRF-2).
DE Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidaea;
OC Aplysiidae; Aplysia.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 87132918.
RX TAUSSIG R., SCHELLER R.H.;
RA "The Aplysia EMRFamide gene encodes sequences related to mammalian
RT brain peptides.";
RN DNA 5:453-461(1986).
RN [2]
RN SEQUENCE OF 87-267 FROM N.A.
RX MEDLINE: 85176970.
RX SCHAEFER M., PICCIOTTO M.R., KREINER T., KALDANY R.R., TAUSSIG R.,
RA SCHELLER R.H.;
RN "Aplysia neurons express a gene encoding multiple EMRFamide
RT neuropeptides.";
RN Cell 41:457-467(1985).
CC -1- FUNCTION: EMRFAMIDE INDUCES CONTRACTIONS IN VISCERAL AND SOMATIC
CC MUSCULATURE AS WELL AS IN THE HEART.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT EMRFAMIDE PRECURSORS ARE
CC GENERATED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE FARP (EMRFAMIDE RELATED PEPTIDE)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14958; AAA27755.1; -;
DR EMBL; M11283; IAB59169.1; -;
DR PFAM; PF01581; FARP; 29.
KW Amidation; Cleavage on pair of basic residues; Repeat;
KW Neuropeptide; Alternative splicing.
FT PEPTIDE 24 27 EMRFAMIDE.
FT PEPTIDE 108 111 EMRFAMIDE.
FT PEPTIDE 115 118 EMRFAMIDE.
FT PEPTIDE 153 156 EMRFAMIDE.
FT PEPTIDE 160 163 EMRFAMIDE.
FT PEPTIDE 175 178 EMRFAMIDE.
FT PEPTIDE 191 194 EMRFAMIDE.
FT PEPTIDE 198 201 EMRFAMIDE.
FT PEPTIDE 214 217 EMRFAMIDE.
FT PEPTIDE 229 232 EMRFAMIDE.
FT PEPTIDE 244 247 EMRFAMIDE.
FT PEPTIDE 259 262 EMRFAMIDE.
FT PEPTIDE 274 277 EMRFAMIDE.
FT PEPTIDE 289 292 EMRFAMIDE.
FT PEPTIDE 304 307 EMRFAMIDE.
FT PEPTIDE 319 322 EMRFAMIDE.
FT PEPTIDE 334 337 EMRFAMIDE.
FT PEPTIDE 349 352 EMRFAMIDE.
FT PEPTIDE 365 368 EMRFAMIDE.
FT PEPTIDE 381 384 EMRFAMIDE.
FT PEPTIDE 397 400 EMRFAMIDE.
FT PEPTIDE 413 416 EMRFAMIDE.
FT PEPTIDE 429 432 EMRFAMIDE.
FT PEPTIDE 445 448 EMRFAMIDE.
FT PEPTIDE 461 464 EMRFAMIDE.
FT PEPTIDE 477 480 EMRFAMIDE.
FT PEPTIDE 493 496 EMRFAMIDE.
FT PEPTIDE 509 512 EMRFAMIDE.
FT PEPTIDE 516 519 EMRFAMIDE.
FT MOD_RES 27 27
FT MOD_RES 111 111
FT AMIDATION (G-28 PROVIDE AMIDE GROUP).
FT AMIDATION (G-112 PROVIDE AMIDE GROUP).

```
FT MOD_RES 118 AMIDATION (G-119 PROVIDE AMIDE GROUP).
FT MOD_RES 156 AMIDATION (G-157 PROVIDE AMIDE GROUP).
FT MOD_RES 163 AMIDATION (G-164 PROVIDE AMIDE GROUP).
FT MOD_RES 178 AMIDATION (G-179 PROVIDE AMIDE GROUP).
FT MOD_RES 194 AMIDATION (G-195 PROVIDE AMIDE GROUP).
FT MOD_RES 201 AMIDATION (G-202 PROVIDE AMIDE GROUP).
FT MOD_RES 217 AMIDATION (G-218 PROVIDE AMIDE GROUP).
FT MOD_RES 232 AMIDATION (G-233 PROVIDE AMIDE GROUP).
FT MOD_RES 247 AMIDATION (G-248 PROVIDE AMIDE GROUP).
FT MOD_RES 262 AMIDATION (G-263 PROVIDE AMIDE GROUP).
FT MOD_RES 277 AMIDATION (G-278 PROVIDE AMIDE GROUP).
FT MOD_RES 292 AMIDATION (G-293 PROVIDE AMIDE GROUP).
FT MOD_RES 307 AMIDATION (G-308 PROVIDE AMIDE GROUP).
FT MOD_RES 322 AMIDATION (G-323 PROVIDE AMIDE GROUP).
FT MOD_RES 337 AMIDATION (G-338 PROVIDE AMIDE GROUP).
FT MOD_RES 352 AMIDATION (G-353 PROVIDE AMIDE GROUP).
FT MOD_RES 368 AMIDATION (G-369 PROVIDE AMIDE GROUP).
FT MOD_RES 384 AMIDATION (G-385 PROVIDE AMIDE GROUP).
FT MOD_RES 400 AMIDATION (G-401 PROVIDE AMIDE GROUP).
FT MOD_RES 416 AMIDATION (G-417 PROVIDE AMIDE GROUP).
FT MOD_RES 432 AMIDATION (G-433 PROVIDE AMIDE GROUP).
FT MOD_RES 448 AMIDATION (G-449 PROVIDE AMIDE GROUP).
FT MOD_RES 464 AMIDATION (G-465 PROVIDE AMIDE GROUP).
FT MOD_RES 480 AMIDATION (G-481 PROVIDE AMIDE GROUP).
FT MOD_RES 496 AMIDATION (G-497 PROVIDE AMIDE GROUP).
FT MOD_RES 512 AMIDATION (G-513 PROVIDE AMIDE GROUP).
FT MOD_RES 528 AMIDATION (G-520 PROVIDE AMIDE GROUP).
SQ SEQUENCE 552 AA; 64483 MW; 644999 CRC32;

Query Match
Best Local Similarity 24.0%; Pred. No. 3.76e-02;
Matches 24; Conservative 25; Mismatches 45; Indels 6; Gaps 6;

Db 357 GSDE-VDKRFMRFGKSLGTD-D-VNKRFRFGKSLGTD-VNKR 412
QY 38 GSPEPCIIHRGKPGGCHGSEPCIIHRGKPGGCHGSEPCIIHR 97
Db 413 FMRFGSLGTD-VDKRFMRFGKSLGTD-D-VNKRFRFGK 450
QY 98 GKPFSSCHGSEPCIIHRGKPGGCHGSEPCIIHRGKPF 137

RESULT
ID Z022_XENLA STANDARD; PRT; 435 AA.
AC P18745;
DC 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE OOCYTE ZINC FINGER PROTEIN XLCP22.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89005062.
RA KOESTER M., PIELER T., POETING A., KNOEHEL W.;
RT "The finger motif defines a multigene family represented in the
RT maternal mRNA of Xenopus laevis oocytes."
RL EMBO J. 7:1735-1741(1988).
DR PIR; S00833; S00833.
DR HSSP; P08047; LSP2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 12.
DR PFAM; PF00096; zf-C2H2; 12.
KW zinc-finger; Metal-binding; DNA-binding; Repeat.
FT ZN_FING 97 120 C2H2-TYPE.
FT ZN_FING 126 148 C2H2-TYPE.
FT ZN_FING 153 176 C2H2-TYPE.
FT ZN_FING 154 176 C2H2-TYPE.
FT ZN_FING 182 204 C2H2-TYPE.
FT ZN_FING 210 232 C2H2-TYPE.
FT ZN_FING 238 260 C2H2-TYPE.
FT ZN_FING 266 288 C2H2-TYPE.
```

```
FT ZN_FING 294 316 C2H2-TYPE.
FT ZN_FING 322 345 C2H2-TYPE.
FT ZN_FING 351 373 C2H2-TYPE.
FT ZN_FING 379 402 C2H2-TYPE.
FT ZN_FING 408 430 C2H2-TYPE.
SQ SEQUENCE 435 AA; 49416 MW; 2203DA04 CRC32;

Query Match
Best Local Similarity 31.2%; Pred. No. 5.50e-02;
Matches 24; Conservative 15; Mismatches 30; Indels 8; Gaps 8;

Db 122 RKPFS-CSCGKCFRSLDHQTHGKPKCCQCKSKFSVRSFLDHRTHGK 180
QY 63 RGRPFSSCHGSEPCIIHRGKPGG-G-CHGSE-P-CIIHRGKPF-SCHGSEPCIIHRG-K 116
Db 181 PFS-CLECGKCFRSLR 196
QY 117 PFGCHGSEPCIIHRGK 133

RESULT
ID PSTA_DICDI STANDARD; PRT; 1046 AA.
AC P11976;
DC 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE PRESTALK PROTEIN PRECURSOR.
DE PRESTALK PROTEIN PRECURSOR.
GN ECMB.
OS Dictyostellum discoideum (Slime mold).
OC Eukaryota; Dictyostellida; Dictyostellium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2.
RX MEDLINE; 88015608.
RA CECCARELLI A., MCROBBIE S.J., JERMYN K.A., DUFFY K., EARLY A.,
RA WILLIAMS J.G.;
RT "Structural and functional characterization of a Dictyostellium gene
RT encoding a DIF inducible, prestalk-enriched mRNA sequence."
RN Nucleic Acids Res. 15:7463-7476(1987).
RP SEQUENCE OF 1-17 FROM N.A.
RX MEDLINE; 87257883.
RA AYRES K., NEUMAN W., ROWEKAMP W.G., CHUNG S.;
RT "Developmental regulation of Dnase I-hypersensitive sites in
RT Dictyostellum discoideum."
RL Mol. Cell. Biol. 7:1823-1829(1987).
CC -I- INDUCTION: BY THE PUTATIVE STALK-SPECIFIC MORPHOGEN DIF
CC (DIFFERENTIATION INDUCING FACTOR).
CC -I- SIMILARITY: THE REPEATED SEQUENCES ARE HIGHLY HOMOLOGOUS TO A
CC REPEAT OF THE PDE3 MRNA OF DICTYOSTELLUM DISCOIDEUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16345; AAA33216.1; -.
DR PIR; A26838; A26838.
DR HSSP; P00592; 3P2P.
DR DICTYDB; DD03005; ECMB.
DR PFAM; PF00526; S_mold_repeat; 42.
KW Signal; Repeat.
FT SIGNAL 1 18 PROBABLE.
FT CHAIN 19 1046 PRESTALK PROTEIN.
FT DOMAIN 31 1012 41 X 24 AA TANDEM REPEATS, CYS-RICH.
FT REPEAT 31 53 X-1.
FT REPEAT 54 77 X-2.
FT REPEAT 78 101 X-3.
FT REPEAT 102 125 A-1.
FT REPEAT 126 149 A-2.
```



```

FT DISULFID 443 452 BY SIMILARITY.
FT DISULFID 455 470 BY SIMILARITY.
FT DISULFID 473 487 BY SIMILARITY.
FT DISULFID 475 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 508 522 BY SIMILARITY.
FT DISULFID 786 798 BY SIMILARITY.
FT DISULFID 788 805 BY SIMILARITY.
FT DISULFID 807 816 BY SIMILARITY.
FT DISULFID 819 831 BY SIMILARITY.
FT DISULFID 834 846 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 867 877 BY SIMILARITY.
FT DISULFID 880 889 BY SIMILARITY.
FT DISULFID 882 896 BY SIMILARITY.
FT DISULFID 899 908 BY SIMILARITY.
FT DISULFID 911 927 BY SIMILARITY.
FT DISULFID 930 946 BY SIMILARITY.
FT DISULFID 932 957 BY SIMILARITY.
FT DISULFID 959 968 BY SIMILARITY.
FT DISULFID 971 986 BY SIMILARITY.
FT DISULFID 989 1003 BY SIMILARITY.
FT DISULFID 991 1010 BY SIMILARITY.
FT DISULFID 1013 1022 BY SIMILARITY.
FT DISULFID 1025 1038 BY SIMILARITY.
FT DISULFID 1098 1110 BY SIMILARITY.
FT DISULFID 1100 1117 BY SIMILARITY.
FT DISULFID 1119 1128 BY SIMILARITY.
FT DISULFID 1131 1143 BY SIMILARITY.
FT DISULFID 1146 1158 BY SIMILARITY.
FT DISULFID 1148 1165 BY SIMILARITY.
FT DISULFID 1167 1176 BY SIMILARITY.
FT DISULFID 1179 1190 BY SIMILARITY.
FT DISULFID 1193 1193 INTERCHAIN (PROBABLE).
FT DISULFID 1196 1196 INTERCHAIN (PROBABLE).
FT DISULFID 1800 1800 POTENTIAL.
FT CARBOHYD 251 251 POTENTIAL.
FT CARBOHYD 371 371 POTENTIAL.
FT CARBOHYD 1088 1088 POTENTIAL.
FT CARBOHYD 1252 1252 POTENTIAL.
FT CARBOHYD 1311 1311 POTENTIAL.
FT CARBOHYD 1351 1351 POTENTIAL.
FT CARBOHYD 1502 1502 POTENTIAL.
SQ SEQUENCE 1801 AA; 196473 MW; BF26F51C CRC32;

Query Match 8.0%; Score 85; DB 1; Length 1801;
Best Local Similarity 27.6%; Pred. No. 5.50e-02;
Matches 24; Conservative 19; Mismatches 35; Indels 9; Gaps 8;

Db 1089 FTSGRCGPCACPSRARGPTCNFTGCHAG--FGG-RTCSCQLRWGDPGLQGRA 1145
Qy 50 FGGCHGSEPCIIHRGKPS-SCHG-SEPCIIHRGKPFGGCHGSEPC-IIHRGKPFSSCHG 106
Db 1146 CD-CD-PRGIDKPCHRSTGHCSCRPG 1170
Qy 107 SEPCIIHRGKPFGGCH-GSEPCIIHRG 132

RESULT 11
ID FMR1_APLCA STANDARD; PRT; 309 AA.
AC P08020;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE FMRFAMIDE NEUROPEPTIDE PRECURSOR (CLONE FMRF-1) (FRAGMENT).
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea;
OC Aplysiidae; Aplysia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85176970.
RA SCHAEFER M., PICCIOTTO M.R., KREINER T., KALDANY R.-R., TAUSSIG R.,

```

```

RA SCHELLER R.H.;
RT "Aplysia neurons express a gene encoding multiple FMRFamide
neuropeptides.";
RL Cell 41:457-467(1985).
CC -1- FUNCTION: FMRFAMIDE INDUCES CONTRACTIONS IN VISCERAL AND SOMATIC
MUSCULATURE AS WELL AS IN THE HEART.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT FMRFAMIDE PRECURSORS ARE
GENERATED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE FARP (FMRFAMIDE RELATED PEPTIDE)
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11282; AAA27751.1; -.
DR PIR; A23172; A23172.
DR PFAM; PF01581; FARP; 19.
KW Amidation; Cleavage on pair of basic residues; Repeat;
Neuropeptide; Alternative splicing.
FT NON_TER 1 1
FT PEPTIDE 7 10 FMRFAMIDE.
FT PEPTIDE 22 25 FMRFAMIDE.
FT PEPTIDE 38 41 FMRFAMIDE.
FT PEPTIDE 45 48 FMRFAMIDE.
FT PEPTIDE 61 64 FMRFAMIDE.
FT PEPTIDE 76 79 FMRFAMIDE.
FT PEPTIDE 91 94 FMRFAMIDE.
FT PEPTIDE 106 109 FMRFAMIDE.
FT PEPTIDE 122 125 FMRFAMIDE.
FT PEPTIDE 138 141 FMRFAMIDE.
FT PEPTIDE 154 157 FMRFAMIDE.
FT PEPTIDE 170 173 FMRFAMIDE.
FT PEPTIDE 186 189 FMRFAMIDE.
FT PEPTIDE 202 205 FMRFAMIDE.
FT PEPTIDE 218 221 FMRFAMIDE.
FT PEPTIDE 234 237 FMRFAMIDE.
FT PEPTIDE 250 253 FMRFAMIDE.
FT PEPTIDE 266 269 FMRFAMIDE.
FT PEPTIDE 273 276 FMRFAMIDE.
FT MOD_RES 10 10 AMIDATION (G-11 PROVIDE AMIDE GROUP).
FT MOD_RES 25 25 AMIDATION (G-26 PROVIDE AMIDE GROUP).
FT MOD_RES 41 41 AMIDATION (G-42 PROVIDE AMIDE GROUP).
FT MOD_RES 48 48 AMIDATION (G-49 PROVIDE AMIDE GROUP).
FT MOD_RES 64 64 AMIDATION (G-65 PROVIDE AMIDE GROUP).
FT MOD_RES 79 79 AMIDATION (G-80 PROVIDE AMIDE GROUP).
FT MOD_RES 94 94 AMIDATION (G-95 PROVIDE AMIDE GROUP).
FT MOD_RES 109 109 AMIDATION (G-110 PROVIDE AMIDE GROUP).
FT MOD_RES 125 125 AMIDATION (G-126 PROVIDE AMIDE GROUP).
FT MOD_RES 141 141 AMIDATION (G-142 PROVIDE AMIDE GROUP).
FT MOD_RES 157 157 AMIDATION (G-158 PROVIDE AMIDE GROUP).
FT MOD_RES 173 173 AMIDATION (G-174 PROVIDE AMIDE GROUP).
FT MOD_RES 189 189 AMIDATION (G-190 PROVIDE AMIDE GROUP).
FT MOD_RES 205 205 AMIDATION (G-206 PROVIDE AMIDE GROUP).
FT MOD_RES 221 221 AMIDATION (G-222 PROVIDE AMIDE GROUP).
FT MOD_RES 237 237 AMIDATION (G-238 PROVIDE AMIDE GROUP).
FT MOD_RES 253 253 AMIDATION (G-254 PROVIDE AMIDE GROUP).
FT MOD_RES 269 269 AMIDATION (G-270 PROVIDE AMIDE GROUP).
FT MOD_RES 276 276 AMIDATION (G-277 PROVIDE AMIDE GROUP).
SQ SEQUENCE 309 AA; 36213 MW; 60CECD67 CRC32;

Query Match 7.9%; Score 84; DB 1; Length 309;
Best Local Similarity 24.0%; Pred. No. 8.02e-02;
Matches 24; Conservative 25; Mismatches 45; Indels 6; Gaps 6;

Db 114 GSDE-VDKRFMRFGKSVGSDE-VDKRFMRFGKSLGTTD-VDKRFMRFGKSLGTTD-VDKR 169
Qy 38 GSEPCIIHRGKPFGGCHGSEPCIIHRGKPFSSCHGSEPCIIHRGKPFSSCHGSEPCIIHR 97

```

DB 170 FMRFGLSGTDD-VDKRFMRFGKSLGTD-VDKRFMRFGK 207
 QY 98 GKPFSSCHGSEPCIIHRGKPFSGCHGSEPCIIHRGKPFSSR 137

RESULT 12

ID VGNMCPMV STANDARD; PRT; 1046 AA.
 AC P03599; Q84103;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE GENOME POLYPROTEIN M [CONTAINS: COAT PROTEIN VP37; COAT PROTEIN VP23].
 OS Cowpea mosaic virus (CPMV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
 OC Comovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 84057775.
 RA VAN WEZENBEEK P., VERVER J.J., HARMSSEN J., VOS P., VAN KAMMEN A.;
 RT "Primary structure and gene organization of the middle-component RNA
 of Cowpea mosaic virus.";
 RL EMBO J. 2:941-946(1983).
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
 EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF THE TWO COAT
 CC PROTEINS.
 CC -!- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-118 IS THE
 CC INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X00729; CAA25314.1;
 CC PIR; A04210; GNWEC.
 CC HSSP; P23009; 1BMV.
 CC Coat protein; Polyprotein.
 CC CHAIN 460 833
 CC 834 1046
 CC SEQUENCE 1046 AA; 116217 MW; 957EB88C CRC32;

Query Match 7.8%; Score 82; DB 1; Length 1046;
 Best Local Similarity 29.4%; Pred. No. 1.69e-01;
 Matches 15; Conservative 14; Mismatches 20; Indels 2; Gaps 2;

DB 699 RYFRELHF-EVTKMSPIKATVTLLAFGNLSDAFGVESPPIRVQFA 748
 QY 2 QYIKANSKPIGITELGQVYIKANSKF-IGITELSSCHGSEPCIIHRGKPGF 51

RESULT 13

ID NOTCDROME STANDARD; PRT; 2703 AA.
 AC P07207; P04154;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NEUROGENIC LOCUS NOTCH PROTEIN PRECURSOR.
 GN N.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 86079539.
 RA WHARTON K.A., JOHANSEN K.M., XU T., ARTAVANIS-TSAKONAS S.;
 RT "Nucleotide sequence from the neurogenic locus notch implies a gene
 product that shares homology with proteins containing EGF-like
 repeats.";

RL Cell 43:567-581(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE; 87064624.

RA KIDD S., KELLEY M.R., YOUNG M.W.;
 RT "Sequence of the notch locus of Drosophila melanogaster: relationship
 of the encoded protein to mammalian clotting and growth factors.";
 RL Mol. Cell. Biol. 6:3094-3108(1986).
 RN [3]
 RP SEQUENCE OF 2505-2611 FROM N.A.
 RX MEDLINE; 85099329.

RA WHARTON K.A., YEDVOBNICK B., FINNERTY V.G., ARTAVANIS-TSAKONAS S.;
 RT "opa: a novel family of transcribed repeats shared by the Notch locus
 and other developmentally regulated loci in D. melanogaster.";
 RL Cell 40:55-62(1985).
 RN [4]
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE; 87257846.

RA KELLEY M.R., KIDD S., BERG R.L., YOUNG M.W.;
 RT "Restriction of P-element insertions at the Notch locus of Drosophila
 melanogaster.";
 RL Mol. Cell. Biol. 7:1545-1548(1987).
 RN [5]
 RP REVIEW.

RA HARRIS W.A.;
 RT "Many cell types specified by Notch function.";
 RL Curr. Biol. 1:120-122(1991).
 CC -!- FUNCTION: NOTCH PROTEIN IS ESSENTIAL FOR PROPER DIFFERENTIATION OF
 CC ECTODERM.

CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- MISCELLANEOUS: SEPARATION OF NEUROBLASTS FROM THE ECTODERM INTO
 CC THE INNER PART OF EMBRYO IS ONE OF THE FIRST STEPS OF CNS
 CC DEVELOPMENT IN INSECTS, THIS PROCESS IS UNDER CONTROL OF THE
 CC NEUROGENIC GENES.

CC -!- SIMILARITY: HIGH, WITH OTHER NOTCH-TYPE PROTEINS.
 CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M16152; AAB59220.1;
 CC EMBL; M16153; AAB59220.1; JOINED.
 CC EMBL; M16149; AAB59220.1; JOINED.
 CC EMBL; M16150; AAB59220.1; JOINED.
 CC EMBL; M16151; AAB59220.1; JOINED.
 CC EMBL; K03508; AAA28725.1;
 CC EMBL; M13689; AAA28725.1; JOINED.
 CC EMBL; K03507; AAA28725.1; JOINED.
 CC EMBL; M12175; AAA74496.1;
 CC PIR; A24420; A24420.
 CC PIR; A24768; A24768.
 CC PIR; A05267; A05267.
 CC HSSP; P00740; 1IXA.

CC FLYBASE; FBgn0004647; N.
 CC PROSITE; PS00010; ASX-HYDROXYL; 22.
 CC PROSITE; PS00022; EGF_1; 34.
 CC PROSITE; PS01186; EGF_2; 28.
 CC PROSITE; PS01187; EGF_CA; 22.
 CC PFAM; PF00008; EGF; 36.
 CC PFAM; PF00023; ank; 6.
 CC PFAM; PF00066; notch; 3.
 CC Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Signal; Glycoprotein.
 FT SIGNAL 1 44

```

FT CHAIN 45 2703 NEUROGENIC LOCUS NOTCH PROTEIN.
FT DOMAIN 45 1745 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1746 1766 POTENTIAL.
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 1451 36 X EGF-TYPE REPEATS.
FT DOMAIN 58 95 EGF-LIKE 1.
FT DOMAIN 96 136 EGF-LIKE 2.
FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 865 EGF-LIKE 21.
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23.
FT DOMAIN 946 982 EGF-LIKE 24.
FT DOMAIN 984 1020 EGF-LIKE 25.
FT DOMAIN 1022 1058 EGF-LIKE 26.
FT DOMAIN 1060 1096 EGF-LIKE 27.
FT DOMAIN 1098 1134 EGF-LIKE 28.
FT DOMAIN 1136 1181 EGF-LIKE 29.
FT DOMAIN 1183 1219 EGF-LIKE 30.
FT DOMAIN 1221 1257 EGF-LIKE 31.
FT DOMAIN 1259 1295 EGF-LIKE 32.
FT DOMAIN 1297 1335 EGF-LIKE 33.
FT DOMAIN 1337 1373 EGF-LIKE 34.
FT DOMAIN 1375 1412 EGF-LIKE 35.
FT DOMAIN 1415 1451 EGF-LIKE 36.
FT DOMAIN 1475 1593 3 X LIN/NOTCH REPEATS.
FT REPEAT 1475 1513 LIN/NOTCH 1.
FT REPEAT 1514 1553 LIN/NOTCH 2.
FT REPEAT 1554 1593 LIN/NOTCH 3.
FT DOMAIN 1896 2109 6 X ANK MOTIF REPEATS.
FT DOMAIN 2338 2568 POLY-GLN (OPA-REPEAT).
FT DISULFID 62 73 BY SIMILARITY.
FT DISULFID 67 83 BY SIMILARITY.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 100 111 BY SIMILARITY.
FT DISULFID 105 124 BY SIMILARITY.
FT DISULFID 126 135 BY SIMILARITY.
FT DISULFID 143 154 BY SIMILARITY.
FT DISULFID 148 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 181 192 BY SIMILARITY.
FT DISULFID 186 203 BY SIMILARITY.
FT DISULFID 205 214 BY SIMILARITY.
FT DISULFID 221 232 BY SIMILARITY.
FT DISULFID 226 241 BY SIMILARITY.
FT DISULFID 243 252 BY SIMILARITY.
FT DISULFID 259 270 BY SIMILARITY.
FT DISULFID 264 279 BY SIMILARITY.
FT DISULFID 281 290 BY SIMILARITY.
FT DISULFID 297 308 BY SIMILARITY.
FT DISULFID 302 317 BY SIMILARITY.
FT DISULFID 319 328 BY SIMILARITY.
FT DISULFID 335 349 BY SIMILARITY.
FT DISULFID 343 358 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 396 BY SIMILARITY.

```

```

FT DISULFID 398 407 BY SIMILARITY.
FT DISULFID 413 424 BY SIMILARITY.
FT DISULFID 418 435 BY SIMILARITY.
FT DISULFID 437 446 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 459 474 BY SIMILARITY.
FT DISULFID 476 485 BY SIMILARITY.
FT DISULFID 492 503 BY SIMILARITY.
FT DISULFID 497 512 BY SIMILARITY.
FT DISULFID 514 523 BY SIMILARITY.
FT DISULFID 530 541 BY SIMILARITY.
FT DISULFID 535 550 BY SIMILARITY.
FT DISULFID 552 561 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 573 588 BY SIMILARITY.
FT DISULFID 590 599 BY SIMILARITY.
FT DISULFID 606 616 BY SIMILARITY.
FT DISULFID 611 625 BY SIMILARITY.
FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 654 BY SIMILARITY.
FT DISULFID 648 663 BY SIMILARITY.
FT DISULFID 665 674 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 686 701 BY SIMILARITY.
FT DISULFID 703 712 BY SIMILARITY.
FT DISULFID 719 730 BY SIMILARITY.
FT DISULFID 724 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.
FT DISULFID 833 844 BY SIMILARITY.
FT DISULFID 838 853 BY SIMILARITY.

```

... Note: remainder of annotations omitted.

Query Match 7.5%; Score 79; DB 1; Length 2703;

Best Local Similarity 22.4%; Pred. No. 5.08e-01;

Matches 19; Conservative 26; Mismatches 31; Indels 9; Gaps 8;

Db 41 LAFANLPTVGRGTALVAASCTSVGQNGTCTQLNGKTYCACDSHYVDYCE-HRNP 99

QY 28 IGITELS-SCHGSEPCIIHRGKPGGCHGSEPCIIH-RGKPFSSC--H-GSEPCIIHRGK 82

Db 100 CNSIRCONGTCTQVTFNRGPGISC 124

QY 83 PFGG-CHGSEPC-IIHR-GKPFSSC 104

RESULT 14 ID SAPE_SARPE STANDARD; PRT; 94 AA.

AC P18313;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE SAPECIN PRECURSOR.

OS Sarcophaga peregrina (Flesh fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Oestroidea; Sarcophagidae; Sarcophaga.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89034216.

RA MATSUYAMA K., NATORI S.;

RT "Molecular cloning of cDNA for sapecin and unique expression of the

sapecin gene during the development of Sarcophaga peregrina.";

RL J. Biol. Chem. 263:17117-17121(1988).

RN [2]

RP SEQUENCE OF 55-94.

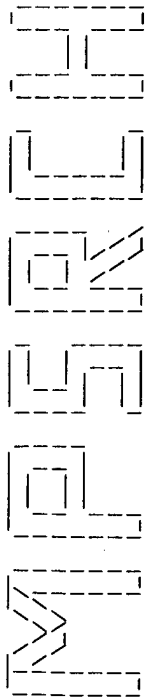
RX MEDLINE; 89034215.

RA MATSUYAMA K., NATORI S.;
 RT "Purification of three antibacterial proteins from the culture medium
 of NIH-Sape-4, an embryonic cell line of *Sarcophaga peregrina*.";
 RL J. Biol. Chem. 263:17112-17116(1988).
 RN [3]
 RN DISULFIDE BONDS.
 RP MEDLINE: 90292974.
 RA KUZUHARA T., NAKAJIMA Y., MATSUYAMA K., NATORI S.;
 RT "Determination of the disulfide array in sapecin, an antibacterial
 peptide of *Sarcophaga peregrina* (flesh fly).";
 RL J. Biochem. 107:514-518(1990).
 RN [4]
 RN STRUCTURE BY NMR.
 RP MEDLINE: 90382590.
 RA HANAWA H., SHIMADA I., KUZUHARA T., KOMANO H., KOHDA D., INAGAKI F.,
 RA NATORI S., ARATA Y.;
 RT "1H nuclear magnetic resonance study of the solution conformation of
 an antibacterial protein, sapecin.";
 RL FEBS Lett. 269:413-420(1990).
 RN [4]
 CC -1- FUNCTION: SAPECINS, WHICH ARE POTENT BACTERICIDAL PROTEINS,
 CC ARE PRODUCED IN RESPONSE TO INJURY. SAPECIN IS CYTOTOXIC TO
 CC GRAM-POSITIVE BACTERIA, AND IN A LESSER EXTENT AGAINST GRAM-
 CC NEGATIVE BACTERIA.
 CC -1- TISSUE SPECIFICITY: HEMOCYTES AND FAT BODY.
 CC -1- INDUCTION: BY INJURY TO THE LARVAL CELL WALL.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD DEFENSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04053; AAA29984.1; -
 DR PIR: A31791; A31791.
 DR PIR: A31792; A31792.
 DR HSSP: P10891; LICA.
 DR PROSITE: PS00425; ARTHROPOD_DEFENSINS; 1.
 DR PFAM: PF01097; Defensin; 1.
 KW Insect immunity; Antibiotic; Cleavage on pair of basic residues;
 KW Signal.
 FT SIGNAL 1 ?
 FT PROPEP 54
 FT PEPTIDE 55 94 SAPECIN.
 FT DISULFID 57 84
 FT DISULFID 70 90
 FT DISULFID 74 92
 SQ SEQUENCE 94 AA; 9914 MW; 6BB5BD33 CRC32;
 Query Match 7.4%; Score 78; DB 1; Length 94;
 Best Local Similarity 28.3%; Pred. No. 7.27e-01;
 Matches 13; Conservative 13; Mismatches 19; Indels 1; Gaps 1;
 DB 47 ELHGRYKATCDLLSGTGNHNSAAHCLL-RGNRGYCNQKAVCV 91
 QY 15 ELGGQYIKANSFIGITELSSCHSGSEPCIHRGKPGGCHGSEPCI 60
 RESULT 15
 ID DEFIN_PROTE STANDARD; PRT; 94 AA.
 AC P10891;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PHORMICIN PRECURSOR (INSECT DEFENSINS A AND B).
 OS *Protophormia terraenovae* (Black blowfly) (Nestling-sucking blowfly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Oestroidea; Calliphoridae; Protophormia.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 90316108.
 RA DIMARCO J.-L., ZACHARY D., HOFFMANN J.A., HOFFMANN D.,
 RA REICHHART J.-M.;
 RT "Insect immunity: expression of the two major inducible antibacterial
 peptides, defensin and dipterizin, in *Phormia terranova*.";
 RL EMBO J. 9:2507-2515(1990).
 RN [2]
 RN SEQUENCE OF 55-94.
 RP TISSUE-HEMOLYMPH;
 RX MEDLINE: 89098894.
 RA LAMBERT J., KEPPI E., DIMARCO J.-L., WICKER C., REICHHART J.-M.,
 RA DUNBAR B., LEPAGE P., VAN DORSSELAER A., HOFFMANN J.A., FOTHERGILL J.,
 RA HOFFMANN D.;
 RT "Insect immunity: Isolation from immune blood of the dipteran *Phormia*
terranova of two insect antibacterial peptides with sequence
 RT homology to rabbit lung macrophage bactericidal peptides.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:262-266(1989).
 RN [3]
 RP DISULFIDE BONDS.
 RX MEDLINE: 91192047.
 RA LEPAGE P., BITSCH F., ROECKLIN D., KEPPI E., DIMARCO J.-L.,
 RA REICHHART J.-M., HOFFMANN J.A., ROITSCH C., VAN DORSSELAER A.;
 RT "Determination of disulfide bridges in natural and recombinant insect
 RT defensin A.";
 RL Eur. J. Biochem. 196:735-742(1991).
 RN [4]
 RN STRUCTURE BY NMR.
 RP MEDLINE: 93104264.
 RA BONMATIN J.M., GENEST M., PETIT M.-C., GINCEL E., SIMORRE J.-P.,
 RA CORNET B., GALLET X., CAILLE A., LABBE H., VOVELLE F., PTAK M.;
 RT "Progress in multidimensional NMR investigations of peptide and
 RT protein 3-D structures in solution. From structure to functional
 RT aspects.";
 RL Biochimie 74:825-836(1992).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE: 95393015.
 RA CORNET B., BONMATIN J.M., HETRU C., HOFFMANN J.A., PTAK M.,
 RA VOVELLE F.;
 RT "Refined three-dimensional solution structure of insect defensin A.";
 RL Structure 3:435-448(1995).
 CC -1- FUNCTION: RESPONSIBLE FOR THE ANTI GRAM-POSITIVE ACTIVITY OF
 CC IMMUNE HEMOLYMPH OF P.TERRANOVAE.
 CC -1- SIMILARITY: BELONGS TO THE ARTHROPOD DEFENSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X55546; CAA39152.1; -
 DR PIR: A32219; A32219.
 DR PIR: B32219; B32219.
 DR PIR: S12558; S12558.
 DR PDB: LICA; 10-JUL-95.
 DR PROSITE: PS00425; ARTHROPOD_DEFENSINS; 1.
 DR PFAM: PF01097; Defensin; 1.
 KW Insect immunity; Antibiotic; Signal; 3D-structure.
 FT SIGNAL 1 23
 FT PROPEP 24 54
 FT CHAIN 55 94 PHORMICIN.
 FT DISULFID 57 84
 FT DISULFID 70 90
 FT DISULFID 74 92
 FT VARIANT 85 86 G -> R (IN DEFENSIN B).
 SQ SEQUENCE 94 AA; 10110 MW; 82F6881A CRC32;
 Query Match 7.4%; Score 78; DB 1; Length 94;
 Best Local Similarity 23.3%; Pred. No. 7.27e-01;
 Matches 14; Conservative 18; Mismatches 26; Indels 2; Gaps 2;

Db 33 HFVDGVOALKEIEPELHGRYKRATCDLLSGTGINHSACAAHCLL-RGNRGYCNKGVCV 91
QY 2 QYIKANSKFIGI-TELGQYIKANSKFIGITELSSCHGSEPCIIHRGKPFGGCHGSEPCI 60

Search completed: Sun Apr 2 17:57:52 2000
Job time : 36 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:56:05 2000; MasPar time 49.78 Seconds

Tabular output not generated.
129.811 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.pep
Perfect Score: 1057
Sequence: 1 DQYKANSKFIGITELGGY.....FGCHGSEPCIIHRGKPFSS 137

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 34.846; Variance 64.388; scale 0.541

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	133	12.6	129	2	JU0394 allergen Der f II (pF)	3.58e-09
2	133	12.6	129	2	A61501 allergen Der f II - h	3.58e-09
3	133	12.6	138	2	B61241 allergen Der f II pre	3.58e-09
4	133	12.6	138	2	A61241 allergen Der f II pre	3.58e-09
5	132	12.5	146	2	A60381 allergen Der p II pre	5.38e-09
6	104	9.8	1315	1	BTCLTN tentoxylisin (EC 3.4.	3.17e-04
7	91	8.6	336	2	S06573 finger protein (clone	3.54e-02
8	90	8.5	1000	2	F70368 cation efflux (AcR8/A	5.02e-02
9	89	8.4	1798	2	S53869 laminin beta-2 chain	7.11e-02
10	88	8.3	313	2	S44208 extracellular matrix	1.00e-01
11	88	8.3	379	2	S50125 larval glue protein L	1.00e-01
12	88	8.3	713	2	JC2534 RVLG protein - rat	1.00e-01
13	87	8.2	597	2	A25790 FMRFamide neuropeptid	1.42e-01
14	85	8.0	435	2	S00833 finger protein (clone	2.79e-01
15	85	8.0	1046	2	A26838 prestalk protein prec	2.79e-01
16	85	8.0	1801	1	MMRTS laminin beta-2 chain	2.79e-01
17	84	7.9	309	2	A23172 FMRFamide neuropeptid	3.91e-01
18	83	7.9	595	2	T04438 hypothetical protein	5.47e-01
19	83	7.9	1797	2	A55677 laminin beta-2 chain	5.47e-01
20	84	7.9	3623	2	T08618 intrinsic factor-B12	3.91e-01
21	82	7.8	1046	1	GNWSC genome polyprotein M	7.62e-01
22	81	7.7	177	2	A24316 eggshell protein - fl	1.06e-00
23	81	7.7	3623	2	T09456 intrinsic factor-B12	1.06e-00

24	78	7.4	94	2	A31792 sapecin precursor - f	2.80e+00
25	78	7.4	94	2	S12558 phormicin A precursor	2.80e+00
26	78	7.4	177	2	A31204 eggshell protein - fl	2.80e+00
27	78	7.4	364	1	D43675 thymidine kinase (EC	2.80e+00
28	78	7.4	637	2	I49638 Drosophila vasa homol	2.80e+00
29	78	7.4	732	1	JG4655 acylaminoacyl-peptida	2.80e+00
30	78	7.4	1042	2	A57534 mucin (clone L31) - h	2.80e+00
31	78	7.4	1220	2	A56136 jagged protein precu	2.80e+00
32	77	7.3	199	2	E71718 hypothetical protein	3.86e+00
33	77	7.3	261	2	S77519 hypothetical protein	3.86e+00
34	77	7.3	1764	2	T15171 hypothetical protein	3.86e+00
35	77	7.3	2517	2	S58380 probable RNA-directed	3.86e+00
36	77	7.3	3566	1	A40701 tenascin-X precursor	3.86e+00
37	76	7.2	107	2	T04306 RIRib protein - rice	5.30e+00
38	76	7.2	107	2	T04305 RIRib protein - rice	5.30e+00
39	76	7.2	123	2	G48677 Ig heavy chain V-D-J	5.30e+00
40	76	7.2	154	2	T06396 isoprenylated protein	5.30e+00
41	76	7.2	755	2	A44315 cartilage oligomeric	5.30e+00
42	75	7.1	200	2	A26637 neurogenic repetitive	7.25e+00
43	74	7.0	313	2	T04776 hypothetical protein	9.89e+00
44	74	7.0	453	2	T05896 pyridoxal-phosphate-d	9.89e+00
45	74	7.0	4006	2	T09070 probable tenascin X -	9.89e+00

ALIGNMENTS

RESULT 1
ENTRY JU0394 #type complete
TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides farinae)
ORGANISM #formal_name Dermatophagoides farinae
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 17-Mar-1999
ACCESSIONS JU0394
REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
#journal Haiga, M.; Okudaira, H.
#title Agric. Biol. Chem. (1991) 55:1233-1238
#cross-references MIM:91291341
#accession JU0394
#molecule_type mRNA
#residues 1-129 #label YUU
CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 12.6%; Score 133; DB 2; Length 129;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSDPCIIHRGKPF 36
|||||:|||||:|||||:
QY 52 GCHGSEPCIIHRGKPF 68

RESULT 2
ENTRY A61501 #type fragment
TITLE allergen Der f II - house-dust mite (Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 13-Sep-1998
ACCESSIONS A61501
REFERENCE A61501
#authors Trudinger, M.; Chua, K.Y.; Thomas, W.R.
#journal Clin. Exp. Allergy (1991) 21:33-37
#title CDNA encoding the major mite allergen Der f II.
#cross-references MIM:91215495
#accession A61501
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-129 #label TRU

```

CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #checksum 476

Query Match 12.6%; Score 133; DB 2; Length 129;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSDPCIHRGKPT 36
|||||:|||||:
QY 52 GCHGSEPCIHRGKPT 68

RESULT 3
ENTRY #type fragment
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides farinae) (fragment)
DATE #formal_name Dermatophagoides farinae
12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS B61241
REFERENCE JU0395
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der p II.
#cross-references MUID:92040281
#accession B61241
##molecule_type mRNA
##residues 1-138 #label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
#domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der p II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 12.6%; Score 133; DB 2; Length 138;
Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 29 GCHGSDPCIHRGKPT 45
|||||:|||||:
QY 52 GCHGSEPCIHRGKPT 68

RESULT 4
ENTRY #type fragment
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides farinae) (fragment)
DATE #formal_name Dermatophagoides farinae
12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS A61241
REFERENCE PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der p II.
#cross-references MUID:92040281
#accession A61241
##molecule_type mRNA
##residues 1-138 #label YUU
##note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
#superfamily allergen Der p II
CLASSIFICATION #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der p II #status experimental #label
MAT
SUMMARY #length 138 #checksum 2894

Query Match 12.6%; Score 133; DB 2; Length 138;

```

```

Best Local Similarity 88.2%; Pred. No. 3.58e-09;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 29 GCHGSDPCIHRGKPT 45
|||||:|||||:
QY 52 GCHGSEPCIHRGKPT 68

RESULT 5
ENTRY #type complete
TITLE allergen Der p II precursor - house-dust mite
ORGANISM (Dermatophagoides pteronyssinus)
DATE #formal_name Dermatophagoides pteronyssinus
03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A60381
REFERENCE A60381
#authors Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;
Stewart, G.A.; Thomas, W.R.
#journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
#title Isolation of cDNA coding for the major mite allergen Der p II
by IgE plaque immunoassay.
#cross-references MUID:90256301
#accession A60381
##status not compared with conceptual translation
##molecule_type mRNA
##residues 1-146 #label CHU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-17
#domain signal sequence #status predicted #label SIG\
#product allergen Der p II #status predicted #label MAT
SUMMARY #length 146 #molecular-weight 15999 #checksum 25

Query Match 12.5%; Score 132; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 5.38e-09;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 37 GCHGSEPCIHRGKPT 52
|||||:|||||:
QY 52 GCHGSEPCIHRGKPT 67

RESULT 6
ENTRY #type complete
TITLE tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani
ALTERNATE_NAMES tetanus neurotoxin
ORGANISM #formal_name Clostridium tetani
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
18-Jun-1999
ACCESSIONS A25689; A25757; A25194; B25194; A60759; S69348; S09364
REFERENCE A25689
#authors Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels,
J.; Weiler, U.; Hudel, M.; Habermann, E.; Niemann, H.
#journal EMBO J. (1986) 5:2495-2502
#title Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
##molecule_type DNA
##residues 1-1315 #label EIS
##cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
##molecule_type DNA
##residues 1-1315 #label FAI
##cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.;

```

```

experimental #label TTH\
: #domain channel forming (fragment B) #status predicted
#label TXB\
#domain ganglioside binding (fragment C) #status
predicted #label TXC\
#binding_site zinc (His) #status predicted\
#active_site Glu #status predicted
#length 1315 #molecular-weight 150681 #checksum 4853

Query Match          9.8%; Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 3.17e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 QYIKANSFIGITEL 844
|||||
QY 2 QYIKANSFIGITEL 16
|||||

RESULT 7
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references
#accession
#status
#molecule_type
#residues
KEYWORDS
SUMMARY

S06573 #type fragment
finger protein (clone Xlcf48-2) - African clawed frog
(fragment)
#formal_name Xenopus laevis #common_name African clawed frog
28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
31-Dec-1993
S06573
S06632
#authors
Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.;
Koster, M.; Poeting, A.; Knoechel, W.
J. Mol. Biol. (1989) 208:639-659
#title
Second-order repeats in Xenopus laevis finger proteins.
#cross-references
MUID:90040698
#accession
S06573
#status
not compared with conceptual translation
#molecule_type mRNA
#residues 1-336 #label NIE
KEYWORDS
DNA binding; zinc finger
SUMMARY
#length 336 #checksum 1745

Query Match          8.6%; Score 91; DB 2; Length 336;
Best Local Similarity 36.7%; Pred. No. 3.54e-02;
Matches 22; Conservative 9; Mismatches 24; Indels 5; Gaps 5;

Db 101 KFSRLTEHRKCHTGEPEFVCSGKSFQSPQLDLMRIHTGKPFCS-CDGCKGCTORA 159
|||||
QY 26 KFIGITELSSCH-GSEPCI-IHRGPFGGCHGSEPCI-IHRG-KPFSSCHGSEPCIHRG 81
|||||

RESULT 8
ENTRY
TITLE
ORGANISM
DATE
ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references
#accession
#status
#molecule_type
#residues
KEYWORDS
SUMMARY

F70368 #type complete
cation efflux (ACrB/ACrD/ACrF family) - Aquifex aeolicus
#formal_name Aquifex aeolicus
08-May-1998 #sequence_revision 08-May-1998 #text_change
08-May-1998
F70368
A70300
#authors
Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.;
Keller, M.; Aujaay, M.; Huber, R.; Feldman, R.A.; Short,
J.M.; Olson, G.J.; Swanson, R.V.
Nature (1998) 392:353-358
#title
The complete genome of the hyperthermophilic bacterium
Aquifex aeolicus.
#cross-references
MUID:98196666
#accession
F70368
#status
preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-1000 #label AQF
#cross-references
GB:AE000707;
#experimental_source strain V59

acrid4
#gene

```

ORGANISM	(fragment)
DATE	#formal_name Dictyostelium minutum
	07-Oct-1994 #sequence_revision 26-Jul-1996 #text_change
	10-Sep-1997
ACCESSIONS	S44208
REFERENCE	S44208
#authors	van Es, S.; Nieuwenhuijsen, B.W.; Lenouvel, F.; van Deursen, E.M.; Schaap, P.
#description	submitted to the EMBL Data Library, April 1994
#accession	Universal signals control slime mold stalk information.
	#molecule_type DNA
	##residues 1-313 ##label VAN
	##cross-references EMBL:X78948; NID:g474918; PID:g474919
GENETICS	
#introns	24/1
KEYWORDS	extracellular matrix
SUMMARY	#length 313 #checksum 3212
	Query Match 8.3%; Score 88; DB 2; Length 313;
	Best Local Similarity 21.9%; Pred. No. 1,00e-01;
	Matches 21; Conservative 18; Mismatches 51; Indels 6; Gaps 5;
DB	84 NCDDNNACTLSDCKKGGCIHTPMGCDHRNCTLPSCDCKGGCT-H--TPMN-CDDKNAC 139
	: ::: :: :: :: :: :: :: ::
QY	35 SCHGSEPCIHRGKPPGGC-HGSEPCIIHRGKPFSSCHGSEPCIHRGKPFSGCHGSEPC 93
	: ::: :: :: :: :: :: :: ::
DB	140 TVDSCNSNPGCSHTPISCDYNSCIVSDCNCGGCV 175
	: ::: :: :: :: :: :: :: ::
QY	94 IHRGKPFSC-HGSEPCIIHRGKPFSGCHGSEPCI 128
	: ::: :: :: :: :: :: :: ::
RESULT	11
ENTRY	S50125 #type complete
TITLE	larval glue protein Lgp3 precursor - fruit fly (Drosophila virilis)
ORGANISM	#formal_name Drosophila virilis
DATE	14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change
	20-Aug-1999
ACCESSIONS	S50125; B60095; S44059
REFERENCE	S50125
#authors	Lanlo, W.; Swida, U.; Kress, H.
#journal	Biochim. Biophys. Acta (1994) 1219:576-580
#title	Molecular cloning of the Drosophila virilis larval glue protein gene Lgp-3 and its comparative analysis with other Drosophila glue protein genes.
	#cross-references MUID:95002181
	#accession S50125
	##molecule_type DNA
	##residues 1-379 ##label LAN
	##cross-references EMBL:Z29565; NID:g450901; PIDN:CAA82671.1; PID:g450902
REFERENCE	A60095
#authors	Swida, U.; Lucka, L.; Kress, H.
#journal	Development (1990) 108:269-280
#title	Glue protein genes in Drosophila virilis: their organization developmental control of transcription and specific mRNA degradation.
	#cross-references MUID:90276249
	#accession B60095
	##molecule_type DNA
	##residues 281-379 ##label SWI
	##cross-references GB:X76203; NID:g433481; PIDN:CAA53797.1; PID:e92940; PID:g1335699
GENETICS	
#gene	FlyBase:Lgp3; Lgp-3
#map position	X16A
#introns	10/1
FUNCTION	
#description	fixes the pupal case to a substrate
CLASSIFICATION	#superfamily salivary glue protein
KEYWORDS	glycoprotein; salivary gland; tandem repeat
FEATURE	

[illegible]

```
##residues      1-1046 #label CEC
##experimental_source strain Ax-2

FEATURE
1-18          #domain signal sequence #status predicted #label SIG\
19-1046      #product prestalk protein #status predicted #label MAT
SUMMARY      #length 1046 #molecular-weight 108823 #checksum 3743

Query Match      8.0%; Score 85; DB 2; Length 1046;
Best Local Similarity 18.8%; Pred. No. 2.79e-01;
Matches 18; Conservative 28; Mismatches 44; Indels 6; Gaps 5;

Db 121 SCDDNNPCTVDSGNSGTGCHTPIINVVHNACTEDKC--IQLGVVTH-TPIA-CDDKNAC 176
QY 35 SCHGSEPCIHR-GKPFGGCHGSEPCIHRGKPFSSCHGSEPCIHRGKPFGGCHGSEPC 93

Db 177 TVDSCSNSTGCCHTPLSCDDNNPCTVDSGNSGTGV 212
QY 94 IHR-GKPFSSCHGSEPCIHRGKPFGGCHGSEPCI 128
```

Search completed: Sun Apr 2 17:56:59 2000
Job time : 54 secs.

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:59:49 2000; MasPar time 3.75 Seconds
Tabular output not generated. 472.933 Million cell updates/sec

Title: >US-09-362-731-3
Description: (1-137) from US09362731.pep
Perfect Score: 1057
Sequence: 1 DOYIKANSKFIGITELGGY.....FGCHGSEPCIIHRGKPFPSR 137

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 23.504; Variance 96.885; scale 0.243

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	134	12.7	61	2	US-08-478- Sequence 162, Applicat	2.05e-04
2	134	12.7	61	2	US-08-482- Sequence 162, Applicat	2.05e-04
3	134	12.7	84	2	US-08-478- Sequence 161, Applicat	2.05e-04
4	134	12.7	84	2	US-08-482- Sequence 161, Applicat	2.05e-04
5	134	12.7	92	2	US-08-478- Sequence 160, Applicat	2.05e-04
6	134	12.7	92	2	US-08-482- Sequence 160, Applicat	2.05e-04
7	134	12.7	129	2	US-08-478- Sequence 157, Applicat	2.05e-04
8	134	12.7	129	2	US-08-478- Sequence 159, Applicat	2.05e-04
9	134	12.7	129	2	US-08-482- Sequence 159, Applicat	2.05e-04
10	134	12.7	129	2	US-08-482- Sequence 157, Applicat	2.05e-04
11	134	12.7	130	2	US-08-478- Sequence 158, Applicat	2.05e-04
12	134	12.7	130	2	US-08-482- Sequence 158, Applicat	2.05e-04
13	133	12.6	129	1	US-08-462- Sequence 8, Applicatio	2.52e-04
14	133	12.6	129	3	PCT-US93-0 Sequence 13, Applicati	2.52e-04
15	133	12.6	129	3	PCT-US93-0 Sequence 8, Applicatio	2.52e-04
16	133	12.6	129	1	US-08-461- Sequence 13, Applicati	2.52e-04
17	133	12.6	129	1	US-08-461- Sequence 8, Applicati	2.52e-04
18	133	12.6	129	1	US-08-461- Sequence 13, Applicati	2.52e-04
19	133	12.6	129	1	US-08-461- Sequence 8, Applicatio	2.52e-04
20	133	12.6	129	1	US-07-945- Sequence 8, Applicatio	2.52e-04
21	133	12.6	129	1	US-08-462- Sequence 13, Applicati	2.52e-04
22	133	12.6	129	2	US-08-478- Sequence 8, Applicatio	2.52e-04
23	133	12.6	129	1	US-07-945- Sequence 13, Applicati	2.52e-04

24	133	12.6	129	2	US-08-482- Sequence 8, Applicatio	2.52e-04
25	133	12.6	142	2	US-08-910- Sequence 4, Applicatio	2.52e-04
26	133	12.6	142	1	US-08-288- Sequence 2, Applicatio	2.52e-04
27	133	12.6	142	2	US-08-905- Sequence 4, Applicatio	2.52e-04
28	133	12.6	142	2	US-08-910- Sequence 2, Applicatio	2.52e-04
29	133	12.6	142	1	US-08-288- Sequence 4, Applicatio	2.52e-04
30	133	12.6	142	2	US-08-905- Sequence 2, Applicatio	2.52e-04
31	132	12.5	25	2	US-08-482- Sequence 42, Applicati	3.11e-04
32	132	12.5	25	3	PCT-US95-0 Sequence 19, Applicati	3.11e-04
33	132	12.5	25	2	US-08-478- Sequence 42, Applicati	3.11e-04
34	132	12.5	129	1	US-08-462- Sequence 12, Applicati	3.11e-04
35	132	12.5	129	1	US-07-945- Sequence 12, Applicati	3.11e-04
36	132	12.5	129	3	PCT-US93-0 Sequence 12, Applicati	3.11e-04
37	132	12.5	129	1	US-08-461- Sequence 12, Applicati	3.11e-04
38	132	12.5	129	1	US-08-461- Sequence 12, Applicati	3.11e-04
39	132	12.5	146	1	US-08-462- Sequence 4, Applicatio	3.11e-04
40	132	12.5	146	1	US-07-945- Sequence 4, Applicatio	3.11e-04
41	132	12.5	146	1	US-08-461- Sequence 4, Applicatio	3.11e-04
42	132	12.5	146	2	US-08-478- Sequence 4, Applicatio	3.11e-04
43	132	12.5	146	3	PCT-US93-0 Sequence 4, Applicatio	3.11e-04
44	132	12.5	146	2	US-08-482- Sequence 4, Applicatio	3.11e-04
45	132	12.5	146	1	US-08-461- Sequence 4, Applicatio	3.11e-04

ALIGNMENTS

RESULT 1
ID US-08-478-572-162 STANDARD; PRT; 61 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 162, Application US/08478572
XX
CC Sequence 162, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 61 AA; 6669 MW; 17328 CN;
Query Match 12.7%; Score 134; DB 2; Length 61;
Best Local Similarity 94.1%; Pred.No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
|||||
QY 52 GCHGSEPCIIHRGKPF 68
|||||
RESULT 2
ID US-08-482-142-162 STANDARD; PRT; 61 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 162, Application US/08482142
CC
CC Sequence 162, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 435
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
Query Match 12.7%; Score 134; DB 2; Length 61;
Best Local Similarity 94.1%; Pred.No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
|||||
QY 52 GCHGSEPCIIHRGKPF 68
|||||
RESULT 3
ID US-08-478-572-161 STANDARD; PRT; 84 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 161, Application US/08478572
CC
CC Sequence 161, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;
Query Match 12.7%; Score 134; DB 2; Length 84;

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 9
ID US-08-482-142-159 STANDARD; PRT; 129 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE
Sequence 159, Application US/08482142
Sequence 159, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA

CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 36
QY 52 GCHGSEPCIIHRGKPF 68
RESULT 10
ID US-08-482-142-157 STANDARD; PRT; 129 AA.
XX AC xxxxxx
XX DT
XX DE
XX DE
Sequence 157, Application US/08482142
Sequence 157, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA

CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14054 MW; 84465 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 129;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCGSEPCIIHRGKPF 36
QY 52 GCGSEPCIIHRGKPF 68
RESULT 11
ID US-08-478-572-158 STANDARD; PRT: 130 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 158, Application US/08478572
XX Sequence 158, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC
CC

CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 130 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 130 AA; 14143 MW; 85527 CN;
SQ
Query Match 12.7%; Score 134; DB 2; Length 130;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 20 GCGSEPCIIHRGKPF 36
QY 52 GCGSEPCIIHRGKPF 68
RESULT 12
ID US-08-482-142-158 STANDARD; PRT: 130 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 158, Application US/08482142
XX Sequence 158, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC

CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 130 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
SQ SEQUENCE 130 AA; 14143 MW; 85527 CN;

Query Match 12.7%; Score 134; DB 2; Length 130;
Best Local Similarity 94.1%; Pred. No. 2.05e-04;
Matches 16; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPT 36
|||||:|||||:
QY 52 GCHGSEPCIIHRGKPTS 68

RESULT 13
ID US-08-462-831-8 STANDARD; PRT; 129 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
Sequence 8, Application US/08462831
Sequence 8, Application US/08462831
Patent No. 5552142
GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/462,831
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 11
CC OTHER INFORMATION: /label-xaa is Asn or Ser
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 52
CC OTHER INFORMATION: /label-xaa is Thr or Ile
CC FEATURE:

CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 129 AA; 14022 MW; 84331 CN;

Query Match 12.6%; Score 133; DB 1; Length 129;
Best Local Similarity 88.2%; Pred. No. 2.52e-04;
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 20 GCHGSDPCIIHRGKPT 36
|||||:|||||:
QY 52 GCHGSEPCIIHRGKPTS 68

RESULT 14
ID PCT-US93-08518-13 STANDARD; PRT; 129 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
Sequence 13, Application PC/TUS9308518
Sequence 13, Application PC/TUS9308518
GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/08518
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/945,288
CC FILING DATE: 10 SEPTEMBER 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 11
CC OTHER INFORMATION: /label-xaa is Asn or Ser
CC FEATURE:
CC NAME/KEY: misc feature
CC LOCATION: 52
CC OTHER INFORMATION: /label-xaa is Thr or Ile
CC FEATURE:

MOLECULE TYPE:

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Result No.	Query Match	Score	Length	DB	ID	Description	Pred. No.
1	117	42.1	343	14	Q82847	HAY SUBUNIT OF HAEMAGG	1.33e-07
2	117	42.1	344	14	Q86890	HAI-HEMAGGLUTININ [GLO	1.33e-07
3	117	42.1	359	14	Q9YJW9	HEMAGGLUTININ PRECURSO	1.33e-07
4	117	42.1	565	14	Q82559	HEMAGGLUTININ.	1.33e-07
5	114	41.0	332	14	Q56964	HAEMAGGLUTININ (FRAGME	4.68e-07
6	114	41.0	335	14	Q91205	HAEMAGGLUTININ (FRAGME	4.68e-07
7	114	41.0	342	14	Q56963	HAEMAGGLUTININ (FRAGME	4.68e-07
8	114	41.0	343	14	Q82848	HAY SUBUNIT OF HAEMAGG	4.68e-07
9	114	41.0	343	14	Q9YJW7	HEMAGGLUTININ PRECURSO	4.68e-07
10	114	41.0	343	14	Q82846	HAY SUBUNIT OF HAEMAGG	4.68e-07
11	114	41.0	343	14	Q82843	HAY SUBUNIT OF HAEMAGG	4.68e-07
12	114	41.0	343	14	Q82844	HAY SUBUNIT OF HAEMAGG	4.68e-07
13	114	41.0	346	14	Q56838	HEMAGGLUTININ HAI SUBU	4.68e-07
14	114	41.0	346	14	Q56857	HEMAGGLUTININ HAI SUBU	4.68e-07
15	114	41.0	353	14	Q12294	HEMAGGLUTININ HAI SUBU	4.68e-07
16	114	41.0	357	14	Q56856	HEMAGGLUTININ HAI SUBU	4.68e-07
17	114	41.0	360	14	Q9YJW3	HEMAGGLUTININ PRECURSO	4.68e-07
18	114	41.0	360	14	Q9YJW8	HEMAGGLUTININ PRECURSO	4.68e-07
19	114	41.0	362	14	Q9YJW9	HEMAGGLUTININ PRECURSO	4.68e-07
20	114	41.0	362	14	Q9YJW6	HEMAGGLUTININ PRECURSO	4.68e-07

```

DR HSP; P03437; 2VIU.
DR PFAM; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTIN12.
DR PRINTS; PR00330; HEMAGGLUTN1.
DR PRINTS; PR00331; HEMAGGLUTN2.
SQ SEQUENCE 565 AA; 63633 MW; A847E68B CRC32;

Query Match          42.1% Score 117; DB 14; Length 565;
Best Local Similarity 63.3%; Pred. No. 1.33e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGK-CPKYIKONTLKLATGMNV 338
    || :: || ||||| |||||::|
Qy 6 QNTLKLATGGKPKYVQNTLKLATGKGK 35

RESULT 5
ID O56964 PRELIMINARY; PRT; 332 AA.
AC O56964;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/BERLIN/1/91(H3N8);
RA ILOBI C.P., NICOLSON C., TAYLOR J., MUMFORD J.A., WOOD J.M.,
RA ROBERTSON J.S.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
REMBL; AJ223196; CAAL1171.1; -
DR HSP; P03437; 2VIU.
DR PFAM; PF00509; Hemagglutinin; 1.
FT NON_TER 1
FT CHAIN 1 >328 HA1 HAEMAGGLUTININ.
FT CHAIN 330 >332 HA2 HAEMAGGLUTININ.
FT NON_TER 332 332
SQ SEQUENCE 332 AA; 36858 MW; C291A548 CRC32;

Query Match          41.0% Score 114; DB 14; Length 332;
Best Local Similarity 60.0%; Pred. No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 295 QNVNKVTYGK-CPKYIKONTLKLATGMNV 323
    || :: || ||||| |||||::|
Qy 6 QNTLKLATGGKPKYVQNTLKLATGKGK 35

RESULT 6
ID O91205 PRELIMINARY; PRT; 335 AA.
AC O91205;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE HAEMAGGLUTININ (FRAGMENT).
GN HA.
OS Influenza A virus.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/EQUINE/BERLIN/4/89(H3N8);
RX MEDLINE; 98309066.
RA ILOBI C.P., NICOLSON C., TAYLOR J., MUMFORD J.A., WOOD J.M.,
RA ROBERTSON J.S.;
RT "Direct sequencing of the HA gene of clinical equine H3N8 influenza
RT virus and comparison with laboratory derived viruses.";
RL Arch. Virol. 143:891-901(1998).
DR EMBL; AJ223195; CAAL1170.1; -.
DR HSP; P03437; 2VIU.
```

Query Match

```

QY      6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35

RESULT 13
ID      056858      PRELIMINARY;      PRT;      346 AA.
AC      O56858;
DT      01-JUN-1998 (TREMBlrel. 06, Created)
DT      01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT      01-JUN-1999 (TREMBlrel. 12, Last annotation update)
DE      DE      HEMAGGLUTININ HAI SUBUNIT PRECURSOR (FRAGMENT).
OS      Equine influenza virus.
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza virus A and B group.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A/EQUI 2/BOLLNAS/96;
RC      OXBURG L., AKERBLUM L., FRIDBERGER T., KLINGEBORN B., LINNE T.;
RL      Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Y14060; CAA74388.1; -
DR      HSSP; P03437; 2VIU.
DR      PFAM; PF00509; Hemagglutinin; 1.
KW      Signal; Hemagglutinin.
FT      SIGNAL      1      15      POTENTIAL.
FT      NON_TER      346      346
SQ      SEQUENCE      346 AA; 38449 MW; 77FE5F34 CRC32;

Query Match      41.0%; Score 114; DB 14; Length 346;
Best Local Similarity 60.0%; Pred. No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db      310 QNVNKVYTGK-CPKYIRONTLKLATGMNV 338
      |||::|||::|::|::|::|::|::|::|
QY      6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35

RESULT 14
ID      056857      PRELIMINARY;      PRT;      346 AA.
AC      O56857;
DT      01-JUN-1998 (TREMBlrel. 06, Created)
DT      01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT      01-JUN-1999 (TREMBlrel. 12, Last annotation update)
DE      DE      HEMAGGLUTININ HAI SUBUNIT PRECURSOR (FRAGMENT).
OS      Equine influenza virus.
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza virus A and B group.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A/EQUI 2/ALVDALEN/96;
RC      OXBURG L., AKERBLUM L., FRIDBERGER T., KLINGEBORN B., LINNE T.;
RL      Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Y14059; CAA74387.1; -
DR      HSSP; P03437; 2VIU.
DR      PFAM; PF00509; Hemagglutinin; 1.
KW      Signal; Hemagglutinin.
FT      SIGNAL      1      15      POTENTIAL.
FT      NON_TER      346      346
SQ      SEQUENCE      346 AA; 38449 MW; 77FE5F34 CRC32;

Query Match      41.0%; Score 114; DB 14; Length 346;
Best Local Similarity 60.0%; Pred. No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db      310 QNVNKVYTGK-CPKYIRONTLKLATGMNV 338
      |||::|||::|::|::|::|::|::|::|
QY      6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35

RESULT 15
ID      012294      PRELIMINARY;      PRT;      353 AA.
AC      012294;
DT      01-JUL-1997 (TREMBlrel. 04, Created)
DT      01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT      01-JUL-1999 (TREMBlrel. 12, Last annotation update)
DE      DE      HEMAGGLUTININ HAI SUBUNIT PRECURSOR (FRAGMENT).
OS      Equine influenza virus.
OC      Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC      Influenza virus A and B group.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=A/EQUI 2/BOLLNAS/96;
RC      OXBURG L., AKERBLUM L., FRIDBERGER T., KLINGEBORN B., LINNE T.;
RL      Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR      EMBL; Y14060; CAA74388.1; -
DR      HSSP; P03437; 2VIU.
DR      PFAM; PF00509; Hemagglutinin; 1.
KW      Signal; Hemagglutinin.
FT      SIGNAL      1      15      POTENTIAL.
FT      NON_TER      346      346
SQ      SEQUENCE      346 AA; 38449 MW; 77FE5F34 CRC32;

Query Match      41.0%; Score 114; DB 14; Length 346;
Best Local Similarity 60.0%; Pred. No. 4.68e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db      310 QNVNKVYTGK-CPKYIRONTLKLATGMNV 338
      |||::|||::|::|::|::|::|::|::|
QY      6 QNTLKLATGKKGPKYVQNTLKLATGKGV 35

```

DE HEMAGGLUTININ HA1 SUBUNIT (FRAGMENT).
 OS Equine influenza virus.
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/EQUI 2/VISINGSO /90;
 RA OXBURGH L.H.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A/EQUI 2/VISINGSO /90;
 RX MEDLINE: 95159661.
 RA OXBURGH L., BERG M., KLINGEBORN B., EMMOTH E., LINNE T.;
 RT "Evolution of H3N8 equine influenza virus from 1963 to 1991."
 RL Virus Res. 34:153-165(1994).
 DR EMBL; Y14056; CAA74384.1; .
 DR HSSP; P03437; 2VIU.
 DR PFAM; PF00509; Hemagglutinin; 1.
 KW Hemagglutinin.
 FT NON_TER 353
 SQ SEQUENCE 353 AA; 39112 MW; 4A3D7FA8 CRC32;

Query Match 41.0%; Score 114; DB 14; Length 353;
 Best Local Similarity 60.0%; Pred.No. 4.68e-07;

Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNVTYTK-CPKYIRONTKLATGMNV 338
 |||::|||::|
 Qy 6 QNTLKLATGKKGPKYKQNTLKLATGKGV 35

Search completed: Sun Apr 2 18:04:44 2000
 Job time : 72 secs.

WIREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:02:45 2000; MasPar time 28.23 Seconds
Tabular output not generated. 42.313 Million cell updates/sec

Title: >US-09-362-731-4
Description: (1-40) from US09362731.pap
Perfect Score: 278
Sequence: 1 PKYVKQNTLKLATGKKGPKYVKQNTLKLATGKKGVIIGIK 40

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:Swissprot

Statistics: Mean 32.905; Variance 55.156; scale 0.597

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	DB	ID	Description	Pred. No.
1	117	42.1	565	1	HEMA_IAHNM HEMAGGLUTININ PRECURSO	2.99e-08
2	117	42.1	565	1	HEMA_IAHFO HEMAGGLUTININ PRECURSO	2.99e-08
3	117	42.1	565	1	HEMA_IAHSA HEMAGGLUTININ PRECURSO	2.99e-08
4	116	41.7	565	1	HEMA_IAHAL HEMAGGLUTININ PRECURSO	2.99e-08
5	115	41.4	565	1	HEMA_IAHUR HEMAGGLUTININ PRECURSO	4.64e-08
6	115	41.4	565	1	HEMA_IAHMI HEMAGGLUTININ PRECURSO	7.19e-08
7	114	41.0	565	1	HEMA_IAHSU HEMAGGLUTININ PRECURSO	1.11e-07
8	114	41.0	565	1	HEMA_IAHK7 HEMAGGLUTININ PRECURSO	1.11e-07
9	114	41.0	565	1	HEMA_IAHK6 HEMAGGLUTININ PRECURSO	1.11e-07
10	114	41.0	565	1	HEMA_IAHTE HEMAGGLUTININ PRECURSO	1.11e-07
11	114	41.0	565	1	HEMA_IAHTO HEMAGGLUTININ PRECURSO	1.11e-07
12	114	41.0	566	1	HEMA_IAUDO HEMAGGLUTININ PRECURSO	1.11e-07
13	110	39.6	328	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
14	110	39.6	328	1	HEMA_IADH6 HEMAGGLUTININ PRECURSO	6.30e-07
15	110	39.6	550	1	HEMA_IADH4 HEMAGGLUTININ PRECURSO	6.30e-07
16	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
17	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
18	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
19	110	39.6	550	1	HEMA_IADH7 HEMAGGLUTININ PRECURSO	6.30e-07
20	110	39.6	550	1	HEMA_IADH3 HEMAGGLUTININ PRECURSO	6.30e-07
21	110	39.6	550	1	HEMA_IADH1 HEMAGGLUTININ PRECURSO	6.30e-07
22	110	39.6	550	1	HEMA_IADH2 HEMAGGLUTININ PRECURSO	6.30e-07
23	110	39.6	550	1	HEMA_IADHM HEMAGGLUTININ PRECURSO	6.30e-07

24	110	39.6	566	1	HEMA_IAME2 HEMAGGLUTININ PRECURSO	6.30e-07
25	110	39.6	566	1	HEMA_IATZC HEMAGGLUTININ PRECURSO	6.30e-07
26	110	39.6	566	1	HEMA_IATAC HEMAGGLUTININ PRECURSO	6.30e-07
27	110	39.6	566	1	HEMA_IADU3 HEMAGGLUTININ PRECURSO	6.30e-07
28	110	39.6	566	1	HEMA_IANT6 HEMAGGLUTININ PRECURSO	6.30e-07
29	110	39.6	566	1	HEMA_IAN7 HEMAGGLUTININ PRECURSO	6.30e-07
30	110	39.6	566	1	HEMA_IATV7 HEMAGGLUTININ PRECURSO	6.30e-07
31	109	39.2	565	1	HEMA_IATRO HEMAGGLUTININ PRECURSO	9.68e-07
32	108	38.8	566	1	HEMA_IATZK HEMAGGLUTININ PRECURSO	1.49e-06
33	107	38.5	550	1	HEMA_IAME6 HEMAGGLUTININ PRECURSO	2.28e-06
34	107	38.5	550	1	HEMA_IADHK HEMAGGLUTININ PRECURSO	2.28e-06
35	107	38.5	550	1	HEMA_IADH3 HEMAGGLUTININ PRECURSO	2.28e-06
36	107	38.5	564	1	HEMA_IABUD HEMAGGLUTININ PRECURSO	2.28e-06
37	106	38.1	566	1	HEMA_IADM2 HEMAGGLUTININ PRECURSO	3.49e-06
38	105	37.8	550	1	HEMA_IAGHK HEMAGGLUTININ PRECURSO	5.32e-06
39	105	37.8	550	1	HEMA_IADHL HEMAGGLUTININ PRECURSO	5.32e-06
40	105	37.8	566	1	HEMA_IADA3 HEMAGGLUTININ PRECURSO	5.32e-06
41	104	37.4	566	1	HEMA_IAMAO HEMAGGLUTININ PRECURSO	8.12e-06
42	103	37.1	564	1	HEMA_IASE2 HEMAGGLUTININ PRECURSO	1.24e-05
43	103	37.1	564	1	HEMA_IATKM HEMAGGLUTININ PRECURSO	1.24e-05
44	103	37.1	564	1	HEMA_IARUD HEMAGGLUTININ PRECURSO	1.24e-05
45	103	37.1	564	1	HEMA_IACKA HEMAGGLUTININ PRECURSO	1.24e-05

ALIGNMENTS

RESULT 1
ID HEMA_IAHNM STANDARD; PRT; 565 AA.
AC P16997; O83997; O83996;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/New Market/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOPRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use, by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M24722; AAA3107.1; ALT_SEQ.
CC FIR; E34064; HMVIES.
CC HSSP; P03437; 1HTM.
CC PFAM; PF00509; Hemagglutinin; 1.
CC Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
FT SEQUENCE 565 AA; 63749 MW; 0E41A00B CRC32;
SQ

```

FT CONFLICT 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63686 MW; 1727DA6B CRC32;

Query Match .. 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 2.99e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGK-CPKYIKQNTLKLATGMNV 338
|||::||| |||:|||||:|||||::|
QY 6 QNTLKLATGKGPKYVQNQLKLATGKKGV 35

RESULT 3
ID HEMA_IAHSA STANDARD; PRT; 565 AA.
AC PL6999; Q84000; Q84001;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
HEMAGGLUTININ HA2 CHAIN].
GN GN
OS Influenza A virus (strain A/Equine/Santiago/1/85).
OC Viruses; SRNA negative-strand viruses; Orthomyxoviridae;
RN Influenza virus A and B group.
RC [1]
RD SEQUENCE FROM N.A.
RE MEDLINE; 89204899.
RX KAWACKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOPRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
(HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24725; AAA43110.1; ALT_SEQ.
DR DR
DR PIR; H34064; HMVE8.
DR HSSP; P03437; LHTM.
DR FRAM; PF00509; Hemagglutinin; Glycoprotein; Signal.
KW Envelope protein; Hemagglutinin;
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63665 MW; FEF4E86F CRC32;

Query Match .. 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 2.99e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKVTYGK-CPKYIKQNTLKLATGMNV 338
|||::||| |||:|||||:|||||::|
QY 6 QNTLKLATGKGPKYVQNQLKLATGKKGV 35

RESULT 4
ID HEMA_IAHAL STANDARD; PRT; 565 AA.
AC PL6994; Q83991; Q83990;
DT 01-AUG-1990 (Rel. 15, Created)

```

DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
 DE HEMAGGLUTININ HA2 CHAIN].
 GN HA.
 OS Influenza A virus (strain A/Equine/Algiers/72).
 OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 89204899.
 RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
 RT Virology 169:283-292(1989).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M24721; AAA43100.1; ALT_SEQ.
 DR PIR; D34064; HMIVB4.
 DR HSSP; P03437; 1HTM.
 DR PFAM; PF00509; Hemagglutinin; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 340 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 23 23 POTENTIAL.
 FT CARBOHYD 37 37 POTENTIAL.
 FT CARBOHYD 53 53 POTENTIAL.
 FT CARBOHYD 78 78 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 300 300 POTENTIAL.
 FT CARBOHYD 498 498 POTENTIAL.
 SQ SEQUENCE 565 AA; 63831 MW; DD483D18 CRC32;
 Query Match 41.7%; Score 116; DB 1; Length 565;
 Best Local Similarity 63.3%; Pred. No. 4.64e-08;
 Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
 DB 310 QNVNKTGYGK-CPKYVKQSTLKLATGMNV 338
 QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35
 RESULT 5
 ID HEMA_IAHUR STANDARD; PRT; 565 AA.
 AC PI7002; O84006; Q84007;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
 DE HEMAGGLUTININ HA2 CHAIN].
 GN HA.
 OS Influenza A virus (strain A/Equine/Uruguay/1/63).
 OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 89204899.
 RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
 RT Virology 169:283-292(1989).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M24718; AAA43114.1; ALT_SEQ.
 DR PIR; A34064; HMIVEL.
 DR HSSP; P03437; JHTM.
 DR PFAM; PF00509; Hemagglutinin; 1.
 KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
 FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
 FT CARBOHYD 37 37 POTENTIAL.
 FT CARBOHYD 53 53 POTENTIAL.
 FT CARBOHYD 78 78 POTENTIAL.
 FT CARBOHYD 180 180 POTENTIAL.
 FT CARBOHYD 300 300 POTENTIAL.
 FT CARBOHYD 498 498 POTENTIAL.
 SQ SEQUENCE 565 AA; 63604 MW; A87356D9 CRC32;
 Query Match 41.4%; Score 115; DB 1; Length 565;
 Best Local Similarity 63.3%; Pred. No. 7.19e-08;
 Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;
 DB 310 QNVNKTGYGK-CPKYVKQSTLKLATGMNV 338
 QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKGV 35
 RESULT 6
 ID HEMA_IAHMI STANDARD; PRT; 565 AA.
 AC PI5658; Q83988; Q83987; Q67097; Q67098;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
 DE HEMAGGLUTININ HA2 CHAIN].
 GN HA.
 OS Influenza A virus (strain A/Equine/Miami/1/63).
 OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group.
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 89204899.
 RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
 RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
 RT Virology 169:283-292(1989).
 [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE; 85133536.
 RA DANIELS R.S., SKEHEL J.J., WILEY D.C.;
 RT "Amino acid sequences of haemagglutinins of influenza viruses of the
 RT H3 subtype isolated from horses.";
 RT J. Gen. Virol. 66:457-464(1985).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
 CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

-----
CC EMBL; M24719; AAA43105.1; ALT_SEQ.
DR EMBL; M29257; AAA43164.1;
DR PIR; B34064; HMIVE2.
DR HSP; P03437; 1HTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
FT CONFLICT 13 15 WVH -> AAD (IN REF. 2).
FT CONFLICT 72 73 RV -> SG (IN REF. 2).
FT CONFLICT 173 173 E -> G (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 377 377 L -> G (IN REF. 2).
FT CONFLICT 380 380 G -> A (IN REF. 2).
FT CONFLICT 400 400 I -> F (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63729 MW; CD915206 CRC32;

Query Match 41.4%; Score 115; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 7, 19e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKKYTK-CPKVKQSTLKLATGMNV 338
|| |:: || |:: ||:: ||:: ||:: ||:: ||
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 7 STANDARD; PRT; 565 AA.
ID HEMA_IAHSU
AC Q08011;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Suffolk/89).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93277383.
RA BINNS M.M., DALY J.M., CHIRNSIDE E.D., MUMFORD J.A., WOOD J.M.,
RT "Genetic and antigenic analysis of an equine influenza H 3 isolate
RT from the 1989 epidemic."
RL Arch. Virol. 130:33-44(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24719; AAA43105.1; ALT_SEQ.
DR PIR; B34065; HMIVE2.
DR HSP; P03437; 1HTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
FT CONFLICT 13 15 WVH -> AAD (IN REF. 2).
FT CONFLICT 72 73 RV -> SG (IN REF. 2).
FT CONFLICT 173 173 E -> G (IN REF. 2).
FT CONFLICT 293 293 V -> W (IN REF. 2).
FT CONFLICT 377 377 L -> G (IN REF. 2).
FT CONFLICT 380 380 G -> A (IN REF. 2).
FT CONFLICT 400 400 I -> F (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
FT CONFLICT 555 555 Q -> E (IN REF. 2).
SQ SEQUENCE 565 AA; 63729 MW; CD915206 CRC32;

Query Match 41.4%; Score 115; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 7, 19e-08;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKKYTK-CPKVKQSTLKLATGMNV 338
|| |:: || |:: ||:: ||:: ||:: ||
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 7 STANDARD; PRT; 565 AA.
ID HEMA_IAHSU
AC Q08011;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Suffolk/89).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93277383.
RA BINNS M.M., DALY J.M., CHIRNSIDE E.D., MUMFORD J.A., WOOD J.M.,
RT "Genetic and antigenic analysis of an equine influenza H 3 isolate
RT from the 1989 epidemic."
RL Arch. Virol. 130:33-44(1993).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24719; AAA43105.1; ALT_SEQ.
DR PIR; B34065; HMIVE2.
DR HSP; P03437; 1HTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

```

```

-----
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN (BY SIMILARITY).
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN (BY SIMILARITY).
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 68 68 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63699 MW; E0A3F55B CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1, 11e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKKYTK-CPKVIQNTLKLATGMNV 338
|| |:: || |:: ||:: ||:: ||:: ||
QY 6 QNTLKLATGKPKYKQNTLKLATGKGV 35

RESULT 8 STANDARD; PRT; 565 AA.
ID HEMA_IAHK7
AC P16996; Q83995; Q83994;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
GN HA.
OS Influenza A virus (strain A/Equine/Kentucky/1/87).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKURA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses."
RL Virology 169:283-292(1989).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24728; AAA43103.1; ALT_SEQ.
DR PIR; B34065; HMIVE2.
DR HSP; P03437; 1HTM.
DR PFAM; PF00509; Hemagglutinin; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 23 23 POTENTIAL.
FT CARBOHYD 37 37 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 68 68 POTENTIAL.
FT CARBOHYD 78 78 POTENTIAL.
FT CARBOHYD 180 180 POTENTIAL.
FT CARBOHYD 300 300 POTENTIAL.
FT CARBOHYD 498 498 POTENTIAL.
SQ SEQUENCE 565 AA; 63702 MW; D963418C CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1, 11e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

```

```

Matches 18: Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMRNV 338
  || ::|| ||::|||::|||::|||::|||
QY 6 QNTLKLATGKGGPKYKQNTLKLATGKGGV 35

RESULT 9
ID HEMA_IAHK6 STANDARD; PRT; 565 AA.
AC P19699;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
HA.
GN Influenza A virus (strain A/Equine/Kentucky/2/86).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC PIR; A34065; HMIVET.
CC HSP; P03437; IHTM.
CC PFAM; PF00509; Hemagglutinin; Glycoprotein; Signal.
CC KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 POTENTIAL.
CC FT CARBOHYD 37 37 POTENTIAL.
CC FT CARBOHYD 53 53 POTENTIAL.
CC FT CARBOHYD 68 68 POTENTIAL.
CC FT CARBOHYD 78 78 POTENTIAL.
CC FT CARBOHYD 180 180 POTENTIAL.
CC FT CARBOHYD 300 300 POTENTIAL.
CC FT CARBOHYD 498 498 POTENTIAL.
CC SQ SEQUENCE 565 AA; 63610 MW; 8114DDB4 CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.11e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMRNV 338
  || ::|| ||::|||::|||::|||::|||
QY 6 QNTLKLATGKGGPKYKQNTLKLATGKGGV 35

RESULT 10
ID HEMA_IAHTE STANDARD; PRT; 565 AA.
AC P17001; Q84004; Q84005;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].

```

```

GN HA.
OS Influenza A virus (strain A/Equine/Tennessee/5/86).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M24726; AAA43112.1; ALT_SEQ.
CC PIR; I34064; HMIVET.
CC HSP; P03437; IHTM.
CC PFAM; PF00509; Hemagglutinin; Glycoprotein; Signal.
CC KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
CC FT SIGNAL 1 16
CC CHAIN 17 343 HEMAGGLUTININ HA1 CHAIN.
CC CHAIN 345 565 HEMAGGLUTININ HA2 CHAIN.
CC FT CARBOHYD 23 23 POTENTIAL.
CC FT CARBOHYD 37 37 POTENTIAL.
CC FT CARBOHYD 53 53 POTENTIAL.
CC FT CARBOHYD 68 68 POTENTIAL.
CC FT CARBOHYD 78 78 POTENTIAL.
CC FT CARBOHYD 180 180 POTENTIAL.
CC FT CARBOHYD 300 300 POTENTIAL.
CC FT CARBOHYD 498 498 POTENTIAL.
CC SQ SEQUENCE 565 AA; 63688 MW; CFF7B83D CRC32;

Query Match 41.0%; Score 114; DB 1; Length 565;
Best Local Similarity 60.0%; Pred. No. 1.11e-07;
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;
Db 310 QNVNKVTYGR-CPKYIRQNTLKLATGMRNV 338
  || ::|| ||::|||::|||::|||::|||
QY 6 QNTLKLATGKGGPKYKQNTLKLATGKGGV 35

RESULT 11
ID HEMA_IAHTO STANDARD; PRT; 565 AA.
AC P17000; Q84002; Q84003;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;
DE HEMAGGLUTININ HA2 CHAIN].
HA.
GN Influenza A virus (strain A/Equine/Tokyo/71).
OS Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89204899.
RA KAWAKA Y., BEAN W.J., WEBSTER R.G.;
RT "Evolution of the hemagglutinin of equine H3 influenza viruses.";
RL Virology 169:283-292(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; M24720; AAA43111.1; ALT_SEQ.

DR PIR; C34064; HMIVE3.

DR HSSP; P03437; LHTM.

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 17 343 HEMAGGLUTININ HA2 CHAIN.

FT CHAIN 345 565 POTENTIAL.

FT CARBOHYD 23 23 POTENTIAL.

FT CARBOHYD 37 37 POTENTIAL.

FT CARBOHYD 53 53 POTENTIAL.

FT CARBOHYD 68 68 POTENTIAL.

FT CARBOHYD 78 78 POTENTIAL.

FT CARBOHYD 180 180 POTENTIAL.

FT CARBOHYD 300 300 POTENTIAL.

FT CARBOHYD 498 498 POTENTIAL.

SQ SEQUENCE 565 AA; 63580 MW; 9A54B199 CRC32;

Query Match 41.08; Score 114; DB 1; Length 565;

Best Local Similarity 60.08; Pred. No. 1.11e-07;

Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKITYGA-CPKYVKQNTLKLATGMNRNI 338

|||:::|||||

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35

RESULT 12

ID HEMA_IAUDO STANDARD; PRT; 566 AA.

AC P19106; Q67022; Q67023;

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HEMAGGLUTININ PRECURSOR [CONTAINS: HEMAGGLUTININ HA1 CHAIN;

DE HEMAGGLUTININ HA2 CHAIN].

GN HA.

OS Influenza A virus (strain A/Udorn/307/72).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group.

CC [1]

RN SEQUENCE FROM N.A.

RP MEDLINE; 85050743.

RA YUFEROV V., KARGINOV V., SAMOKHYALOV E., CHIZHIKOV V., VASILONKO S.,

RA URYVAEV L., ZHDANOV V.M.;

RA RT A/Udorn/307/72 (H3N2).;

RT "Nucleotide sequence of the hemagglutinin gene of the Influenza virus

Do1. Akad. Nauk SSSR 278:738-742(1984).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M54895; AAA43099.1; ..

DR EMBL; M54895; CAB25932.1; ALT_TERM.

DR EMBL; M54895; CAB25933.1; ALT_INIT.

DR HSSP; P03437; LHTM.

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.

FT SIGNAL 1 16

FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.

FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.

FT CARBOHYD 24 24 POTENTIAL.

FT CARBOHYD 38 38 POTENTIAL.

FT CARBOHYD 54 54 POTENTIAL.

FT CARBOHYD 97 97 POTENTIAL.

FT CARBOHYD 181 181 POTENTIAL.

FT CARBOHYD 301 301 POTENTIAL.

FT CARBOHYD 499 499 POTENTIAL.

SQ SEQUENCE 566 AA; 63190 MW; 17EABF51 CRC32;

Query Match 41.08; Score 114; DB 1; Length 566;

Best Local Similarity 63.38; Pred. No. 1.11e-07;

Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 311 QNVNKITYGA-CPKYVKQNTLKLATGMNRN 339

|||:::|||||

QY 6 QNTLKLATGKKGPKYVKQNTLKLATGKKG 35

RESULT 13

ID HEMA_IAQ07 STANDARD; PRT; 328 AA.

AC P04663;

DT 13-AUG-1987 (Rel. 05, Created)

DT 13-AUG-1987 (Rel. 05, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HEMAGGLUTININ [CONTAINS: HEMAGGLUTININ HA1 CHAIN] (FRAGMENT).

GN HA.

OS Influenza A virus (strain A/Qu/7/70).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group.

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE; 81194918.

RA SLEIGH M.J., BOTH G.W., UNDERWOOD P.A., BENDER V.J.;

RT "Antigenic drift in the hemagglutinin of the Hong Kong influenza

RT subtype: correlation of amino acid changes with alterations in viral

RT antigenicity.";

RL J. Virol. 37:845-853(1981).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMER IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; K03338; AAA43195.1; ..

DR PFAM; PF00509; Hemagglutinin; 1.

KW Envelope protein; Hemagglutinin; Glycoprotein.

FT NON_TER 1 1

FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.

FT CARBOHYD 8 8 POTENTIAL.

FT CARBOHYD 22 22 POTENTIAL.

FT CARBOHYD 38 38 POTENTIAL.

FT CARBOHYD 81 81 POTENTIAL.

FT CARBOHYD 165 165 POTENTIAL.

FT CARBOHYD 285 285 POTENTIAL.

FT NON_TER 328 328

SQ SEQUENCE 328 AA; 36059 MW; E6980E2B CRC32;

Query Match 39.68; Score 110; DB 1; Length 328;

Best Local Similarity 63.38; Pred. No. 6.30e-07;

Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 295 QNVNKITYGA-CPKYVKQNTLKLATGMNRN 323

(TM)

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	117	42.1	565	1	HMIVE6	hemagglutinin precurs	5.59e-07
2	117	42.1	565	1	HMIVE8	hemagglutinin precurs	5.59e-07
3	117	42.1	565	1	HMIVE5	hemagglutinin precurs	5.59e-07
4	116	41.7	565	1	HMIVE4	hemagglutinin precurs	8.30e-07
5	115	41.4	565	1	HMIVE1	hemagglutinin precurs	1.23e-06
6	115	41.4	565	1	HMIVE2	hemagglutinin precurs	1.23e-06
7	114	41.0	565	2	S33703	hemagglutinin - influ	1.82e-06
8	114	41.0	565	1	HMIVE7	hemagglutinin precurs	1.82e-06
9	114	41.0	565	1	HMIVE5	hemagglutinin precurs	1.82e-06
10	114	41.0	565	1	HMIVE3	hemagglutinin precurs	1.82e-06
11	114	41.0	565	1	HMIVE9	hemagglutinin precurs	1.82e-06
12	110	39.6	550	2	J01156	hemagglutinin precurs	8.64e-06
13	110	39.6	550	2	HMIV80	hemagglutinin precurs	8.64e-06
14	110	39.6	550	1	HMIV15	hemagglutinin precurs	8.64e-06
15	110	39.6	550	1	HMIV53	hemagglutinin precurs	8.64e-06
16	110	39.6	550	1	HMIV98	hemagglutinin precurs	8.64e-06
17	110	39.6	550	2	A92979	hemagglutinin precurs	8.64e-06
18	110	39.6	550	1	HMIV33	hemagglutinin precurs	8.64e-06
19	110	39.6	550	1	HMIV77	hemagglutinin precurs	8.64e-06
20	110	39.6	550	1	HMIV89	hemagglutinin precurs	8.64e-06
21	110	39.6	566	1	HMIVDU	hemagglutinin precurs	8.64e-06
22	110	39.6	566	1	HMIVHA	hemagglutinin precurs	8.64e-06
23	110	39.6	566	1	HMIVHM	hemagglutinin precurs	8.64e-06

```

CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

FEATURE
1-16 : #domain signal sequence #status predicted #label SIG\
17-344 : #product hemagglutinin HA1 #status predicted #label HA1\
345-565 : #product hemagglutinin HA2 #status predicted #label HA2\
535-551 : #domain transmembrane #status predicted #label TM1\
23,37,53,78,180,
300,498 #binding_site carbohydrate (Asn) (covalent) #status
predicted\

29-481,67-292,
79-91,154-488,
296-320 #disulfide_bonds #status predicted\
554,561,564 #binding_site palmitate (Cys) (covalent) #status
predicted

SUMMARY #length 565 #molecular-weight 63749 #checksum 4207

Query Match 42.1%; Score 117; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 5.59e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKITYGK-CPKYIKONTLKLATGMENV 338
|| ::: || |||:|||||:||||: |
QY 6 QNTLKLATGKRGPYKVKONTLKLATGKRGV 35
18-Jul-1999

ACCESSIONS D34064
REFERENCE A34064
#authors Kawaoaka, Y.; Bean, W.J.; Webster, R.G.
#journal Virology (1989) 169:283-292
#title Evolution of the hemagglutinin of equine H3 influenza viruses.

#cross-references MUID:89204899
#accession D34064
#molecule_type genomic RNA
#residues 1-565 ##label KAW
##cross-references GB:M24721; GB:J04336; NID:g323996; PIDN:AAA43100.1; PID:g323997

GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond

FEATURE
1-16 : #domain signal sequence #status predicted #label SIG\
17-344 : #product hemagglutinin HA1 #status predicted #label HA1\
345-565 : #product hemagglutinin HA2 #status predicted #label HA2\
535-551 : #domain transmembrane #status predicted #label TM1\
23,37,53,78,180,
300,498 #binding_site carbohydrate (Asn) (covalent) #status
predicted\

29-481,67-292,
79-91,154-488,
296-320 #disulfide_bonds #status predicted\
554,561,564 #binding_site palmitate (Cys) (covalent) #status
predicted

SUMMARY #length 565 #molecular-weight 63831 #checksum 4800

Query Match : 41.7%; Score 116; DB 1; Length 565;
Best Local Similarity 63.3%; Pred. No. 8.30e-07;
Matches 19; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Db 310 QNVNKITYGK-CPKYVKGSTLKLATGMENV 338
|| ::: || |||:|||||:||||: |

```

GENETICS

#title	Evolution of the hemagglutinin of equine H3 influenza viruses.
#cross-references MUID:	89204899
#accession A34065	
#molecule_type genomic RNA	
#residues I-565 ##label KAW	
#cros-references GB:M24727; GB:J04336;	NID:g324000; PIDN:AAA43102.1; PID:g324001
GENETICS	
#map_position segment 4	
CLASSIFICATION #superfamily influenza virus hemagglutinin glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond	
FEATURE	
1-16 #domain signal sequence #status predicted #label SIG\	
17-344 #product hemagglutinin HA1 #status predicted #label HA1\	
345-565 #product hemagglutinin HA2 #status predicted #label HA2\	
535-551 #domain transmembrane #status predicted #label TM1\	
18,23,37,53,68,78, 180,300,498 #binding_site carbohydrate (Asn) (covalent) #status predicted\	
29-481,67-292,	
79-91,154-488,	
296-320 #disulfide_bonds #status predicted\	
554,561,564 #binding_site palmitate (Cys) (covalent) #status predicted	
SUMMARY #length 565 #molecular-weight 63610 #checksum 6051	
Query Match 41.0%; Score 114; DB 1; Length 565;	
Best Local Similarity 60.0%; Pred.No.1.82e-06;	
Matches 18; Conservative 6; Mismatches 5; Indels 1; Gaps 1;	
Db 310 QWNVKTYGK-CPKYRQNTLKLATGMVY 338	
I I : : : :	
QY 6 QNTLKLATGGKKPKYVKONTCLKLAIGKKGV 35	
RESULT 9	
ENTRY HMIVEE #type complete	
TITLE hemagglutinin precursor - Influenza A virus (strain A/equine/Kentucky/1/87[H3N8])	
CONTAINS hemagglutinin HA1; hemagglutinin HA2	
ORGANISM #formal_name Influenza A virus	
DATE 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 16-Jul-1999	
ACCESSIONS B34065	
REFERENCE A34064	
#authors Kawakita, Y.; Bean, W.J.; Webster, R.G.	
#journal Virology (1989) 169:283-292	
#title Evolution of the hemagglutinin of equine H3 influenza viruses.	
#cross-references MUID:89204899	
#accession B34065	
#molecule_type genomic RNA	
#residues I-565 ##label KAW	
#cros-references GB:M24728; GB:J04336;	NID:g324002; PIDN:AAA43103.1; PID:g324003
GENETICS	
#map_position segment 4	
CLASSIFICATION #superfamily influenza virus hemagglutinin glycoprotein; hemagglutinin; homotrimer; lipoprotein; thiolester bond	
FEATURE	
1-16 #domain signal sequence #status predicted #label SIG\	
17-344 #product hemagglutinin HA1 #status predicted #label HA1\	
345-565 #product hemagglutinin HA2 #status predicted #label HA2\	
535-551 #domain transmembrane #status predicted #label TM1\	
18,23,37,53,68,78, 180,300,498 #binding site carbohydrate (Asn) (covalent) #status predicted\	
29-481,67-292,	
79-91,154-488,	
296-320 #disulfide_bonds #status predicted\	


```

Best Local Similarity 63.3%; Pred.No.8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db    295 QNVNKITYGA-CPKYVKONTLKTATGMNRV 323
      || : :: | ||||| ||||| :: |
QY    6 QNTLKLTGCKGPKYVKONTLKTATGKGV 35

RESULT 13
ENTRY   #type fragment
ENTRY TITLE hemagglutinin precursor - influenza A virus (strain
           A/duck/Hokkaido/8/80) (fragment)
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name Influenza A virus
DATE     30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
         18-Sep-1998
ACCESSIONS B27813
REFERENCE  A94363
#authors Kida, H.; Kawabata, Y.; Naeve, C.W.; Webster, R.G.
#journal Virology (1987) 159:109-119
#title Antigenic and genetic conservation of H3 influenza virus in
        wild ducks.
#cros-references MUID:87265458
#accession B27813
#molecule_type genomic RNA
#residues 1-550 ##label KID
##cros-references GB:M16738; NID:g324083
##note the translation in Fig. 2 is inconsistent with the
        nucleotide sequence in Fig. 1 in having 137-Asn rather
        than Lys for codon AAG

GENETICS
#map_position segment 4
CLASSIFICATION #superfamily Influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
          thiolester bond

FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1
330-550 #product hemagglutinin HA2 #status predicted #label HA2
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\

14-466,52-277, #disulfide_bonds #status predicted\
64-76,139-473, #binding_site palmitate (Cys) (covalent) #status
281-305 predicted
539,546,549 #length 550 #checksnum 4376

SUMMARY
Query Match 39.6%; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred.No.8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Ddb    295 QNVNKITYGA-CPKYVKONTLKTATGMNRV 323
      || : :: | ||||| ||||| :: |
QY    6 QNTLKLTGCKGPKYVKONTLKTATGKGV 35

RESULT 14
ENTRY   #type fragment
ENTRY TITLE hemagglutinin precursor - influenza A virus (strain
           A/duck/Hokkaido/10/85) (fragment)
CONTAINS hemagglutinin HA1; hemagglutinin HA2
ORGANISM #formal_name Influenza A virus
DATE     30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
         16-Jul-1999
ACCESSIONS B27813
REFERENCE  A94363
#authors Kida, H.; Kawabata, Y.; Naeve, C.W.; Webster, R.G.
#journal Virology (1987) 159:109-119
#title Antigenic and genetic conservation of H3 influenza virus in
        wild ducks.
#cros-references MUID:87265458

```

```
#accession G27813
#molecule_type genomic RNA
##residues 1-550 ##label KID
##cross-references GB:M16743; NID:g324093; PIDN:AAA43149.1; PID:g324094
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
thiolester bond
FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1\
330-550 #product hemagglutinin HA2 #status predicted #label HA2\
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\
14-466,52-277,
64-76,139-473,
281-305 #disulfide_bonds #status predicted\
539,546,549 #binding_site palmitate (Cys) (covalent) #status
predicted
SUMMARY #length 550 #checksum 5685
Query Match 39.68; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 295 QNVNKITYGA-CPKYVKONTLKLATGMNV 323
QY 6 ONTLKLATGKKGPKYVKONTLKLATGKKG 35
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
thiolester bond
FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1\
330-550 #product hemagglutinin HA2 #status predicted #label HA2\
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\
14-466,52-277,
64-76,139-473,
281-305 #disulfide_bonds #status predicted\
539,546,549 #binding_site palmitate (Cys) (covalent) #status
predicted
SUMMARY #length 550 #checksum 3973
Query Match 39.68; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 295 QNVNKITYGA-CPKYVKONTLKLATGMNV 323
QY 6 ONTLKLATGKKGPKYVKONTLKLATGKKG 35
GENETICS
#map_position segment 4
CLASSIFICATION #superfamily influenza virus hemagglutinin
KEYWORDS glycoprotein; hemagglutinin; homotrimer; lipoprotein;
thiolester bond
FEATURE
1-328 #product hemagglutinin HA1 #status predicted #label HA1\
330-550 #product hemagglutinin HA2 #status predicted #label HA2\
520-536 #domain transmembrane #status predicted #label TM1\
8,22,38,165,285, #binding_site carbohydrate (Asn) (covalent) #status
483 predicted\
14-466,52-277,
64-76,139-473,
281-305 #disulfide_bonds #status predicted\
539,546,549 #binding_site palmitate (Cys) (covalent) #status
predicted
SUMMARY #length 550 #checksum 3973
```

```
Query Match 39.68; Score 110; DB 1; Length 550;
Best Local Similarity 63.3%; Pred. No. 8.64e-06;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 295 QNVNKITYGA-CPKYVKONTLKLATGMNV 323
QY 6 ONTLKLATGKKGPKYVKONTLKLATGKKG 35
Search completed: Sun Apr 2 18:02:28 2000
Job time : 48 secs.
```

M I S R L A

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Apr 2 18:08:41 2000; MasPar time 67.05 Seconds
33.089 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QYKANSKFIGHTELGCGHSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sprembl12
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-plant 10:sp-rodent 11:sp-rodent 12:sp-unclassified
13:sp-virus 14:sp-virus

Statistics: Mean 30.267; Variance 37.728; scale 0.802
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	98	40.7	145	5	GROUP 2 ALLERGEN EUR M	4.91e-08
2	75	31.1	533	6	RETINAL PIGMENT EPITHE	1.14e-02
3	75	31.1	533	6	RETINAL PIGMENT EPITHE	1.14e-02
4	75	31.1	533	4	RETINAL PIGMENT EPITHE	1.14e-02
5	75	31.1	533	6	RETINAL PIGMENT EPITHE	1.14e-02
6	75	31.1	533	6	RETINAL PIGMENT EPITHE	1.14e-02
7	75	31.1	533	11	RETINAL PIGMENT EPITHE	1.14e-02
8	72	29.9	173	10	T9J22.4 PROTEIN.	5.00e-02
9	70	29.0	116	10	PUTATIVE RIBOSOMAL PRO	1.31e-01
10	69	28.6	102	10	40S RIBOSOMAL PROTEIN	2.12e-01
11	69	28.6	133	10	F6P22.28 PROTEIN.	2.12e-01
12	69	28.6	502	10	HYPOTHETICAL 55.2 KD P	3.40e-01
13	68	28.2	62	2	UREASE-ENHANCING FACTO	3.40e-01
14	68	28.2	382	2	HOMOSERINE O-ACETYLTRA	3.40e-01
15	68	28.2	1036	1	HYPOTHETICAL 112.4 KD	3.40e-01
16	68	28.2	1067	13	PHOSPHOLIPASE C-GAMMA-	3.40e-01
17	66	27.4	379	2	HOMOSERINE O-ACETYLTRA	8.61e-01
18	65	27.0	131	7	MHC CLASS II BETA CHAI	1.36e+00
19	65	27.0	356	5	PERITROPIN-44 PRECURS	1.36e+00
20	65	27.0	1215	5	MYOSIN-IA.	1.36e+00

21	64	26.6	38	2	O25251	2.14e+00
22	64	26.6	159	1	O9YE65	2.14e+00
23	64	26.6	583	11	O54714	2.14e+00
24	63	26.1	138	10	O9XG80	3.35e+00
25	63	26.1	392	13	O9W641	3.35e+00
26	62	25.7	131	7	O46866	5.21e+00
27	62	25.7	149	11	O9Z0J0	5.21e+00
28	62	25.7	163	2	O9X126	5.21e+00
29	62	25.7	219	1	O28237	5.21e+00
30	62	25.7	445	5	O45870	5.21e+00
31	62	25.7	598	2	O34469	5.21e+00
32	62	25.7	665	10	O49236	5.21e+00
33	62	25.7	899	14	O9YTK4	5.21e+00
34	61	25.3	131	14	O9YPT2	8.06e+00
35	61	25.3	156	1	P95911	8.06e+00
36	61	25.3	157	5	O01698	8.06e+00
37	61	25.3	208	3	O9Y7P0	8.06e+00
38	61	25.3	264	2	O9X7C6	8.06e+00
39	61	25.3	349	3	P78872	8.06e+00
40	61	25.3	361	10	P93257	8.06e+00
41	61	25.3	376	2	O9WVA3	8.06e+00
42	61	25.3	433	2	O9Y748	8.06e+00
43	61	25.3	468	5	O9XV51	8.06e+00
44	61	25.3	667	6	O46480	8.06e+00
45	61	25.3	1388	5	Q18566	8.06e+00

ALIGNMENTS

RESULT 1
ID O96430 PRELIMINARY; PRT; 145 AA.
AC O96430;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE GROUP 2 ALLERGEN EUR M 2 0101.
GN EUR M 2 0101.
OS Eukaryophus maynei (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcophormes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Euroglyphus.
RN [1]
RP SEQUENCE FROM N.A.
RA SMITH W., HART B.J., THOMAS W.R.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF047613; AAC82349.1;
DR HSP; P49278.1A9V.
SQ SEQUENCE 145 AA; 15747 MW; 5EF04F1D CRC32;

Query Match 40.7%; Score 98; DB 5; Length 145;
Best Local Similarity 75.0%; Pred. No. 4.91e-08;
Matches 12; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 36 GKGSEPCVHRGTF 51
QY 17 GKGSEPCVHRGTF 32
|| ||||| ||||| :|

RESULT 2
ID Q9XT71 PRELIMINARY; PRT; 533 AA.
AC Q9XT71;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE RETINAL PIGMENT EPITHELIUM-SPECIFIC PROTEIN RPE65.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
OC Chlorocebus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY;
RA MA J.-X., ZHANG D., BROWNLEE N., RE G.R., HAZEN-MARTIN D.J.,

DR	EMBL; U20476; AAC14586.1; JOINED.
DR	EMBL; U20477; AAC14586.1; JOINED.
DR	EMBL; U20478; AAC14586.1; JOINED.
DR	EMBL; U20479; AAC14586.1; JOINED.
DR	EMBL; U20481; AAC14586.1; JOINED.
DR	EMBL; U20482; AAC14586.1; JOINED.
DR	EMBL; U20484; AAC14586.1; JOINED.
DR	EMBL; U20485; AAC14586.1; JOINED.
DR	EMBL; U20486; AAC14586.1; JOINED.
DR	EMBL; AF039868; AAC39660.1; -
DR	EMBL; AF039855; AAC39660.1; JOINED.
DR	EMBL; AF039856; AAC39660.1; JOINED.
DR	EMBL; AF039857; AAC39660.1; JOINED.
DR	EMBL; AF039858; AAC39660.1; JOINED.
DR	EMBL; AF039859; AAC39660.1; JOINED.
DR	EMBL; AF039860; AAC39660.1; JOINED.
DR	EMBL; AF039861; AAC39660.1; JOINED.
DR	EMBL; AF039862; AAC39660.1; JOINED.
DR	EMBL; AF039863; AAC39660.1; JOINED.
DR	EMBL; AF039864; AAC39660.1; JOINED.
DR	EMBL; AF039865; AAC39660.1; JOINED.
DR	EMBL; AF039866; AAC39660.1; JOINED.
DR	EMBL; AF039867; AAC39660.1; JOINED.
SQ	SEQUENCE 533 AA; 60947 MW; OB8A41C3 CRC32;
Query Match 31.1%; Score 75; DB 4; Length 533;	
Best Local Similarity 44.0%; Pred. No. 1.14e-02;	
Matches' 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;	
DB	89 YVRAMTEKRIVITEFGTCAPDPCK 113 ::: : :
QY	2 YIKANS-KFIGITELGGCHGSEPCN 25 ::: : :
RESULT 5	
ID Q05661 PRELIMINARY; PRT; 533 AA.	
AC Q05661;	
DT 01-NOV-1996 (TREMBLrel. 01, Created)	
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)	
DE RETINAL PIGMENT EPITHELIAL MEMBRANE RECEPTOR P63.	
OS Bos taurus (Bovine).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC Bovinae; Bos.	
RN [1]	
RP SEQUENCE FROM N.A., AND SEQUENCE OF 365-382 AND 407-423.	
RC STRAIN-STERNHOLTZ; TISSUE-EYE;	
RX MEDLINE; 93388633.	
RA BAVIK C.O., HELLMAN U., WERNSTEDT C., ERIKSSON U.;	
RT "The retinal pigment epithelial membrane receptor for plasma retinol-binding protein. Isolation and cDNA cloning of the 63-kDa protein.";	
RT J. Biol. Chem. 268:20540-20546(1993).	
CC CC -!- FUNCTION: THIS RECEPTOR BINDS PLASMA RETINOL BINDING PROTEIN.	
CC CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.	
CC CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN RETINAL PIGMENT EPITHELIUM.	
CC EMBL; X66277; CA446988.1; -	
KW Receptor; Retinol-binding; Plasma; Membrane.	
FT MOD_RES 71 71 BLOCKED.	
SQ SEQUENCE 533 AA; 60970 MW; F77E2D7 CRC32;	
Query Match 31.1%; Score 75; DB 6; Length 533;	
Best Local Similarity 44.0%; Pred. No. 1.14e-02;	
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;	
DB	89 YVRAMTEKRIVITEFGTCAPDPCK 113 ::: : :
QY	2 YIKANS-KFIGITELGGCHGSEPCN 25 ::: : :
RESULT 6	
ID Q28175 PRELIMINARY; PRT; 533 AA.	

```

RESULT      8
O048704     PRELIMINARY;          PRT;       173 AA.
AAC         O48704;
ACD         01-JUN-1998 (TReMBLrel. 06, Created)
DDT         01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DTI         01-AUG-1998 (TReMBLrel. 07, Last annotation update)
DE         T9J22.4 PROTEIN.
DE         T9J22.4
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumolliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
OC [1]
RN SEQUENCE FROM N.A.
RR STRAIN=CV. COLUMBIA;
RC ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSSBY M.L., BRANDON R.C.,
RA SYRES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
RRL SOMERVILLE C.R., VENTER J.C.;
RRL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
EMBL; AC002055; AAC14488.1; -.
DSQ SEQUENCE 173 AA; 19015 MW; 01243A3B CRC32;

Query Match           29.9%; Score 72; DB 10; Length 173;
Best Local Similarity 35.5%; Pred. No. 5,00e-02;
Matches 11; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

Ddb      88 YENVNSTINY-DLSVVMGEDPCSIPEGEKF 117
        | : || - : | : | : ||| : | | |
QY       2 YKANSKFIGITELGGCHGSEPCNIHRGKPF 32

RESULT      9
O09XIA6     PRELIMINARY;          PRT;       116 AA.
AAC         O9XIA6;
ACD         01-NOV-1999 (TReMBLrel. 12, Created)
DDT         01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DTI         01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE PUTATIVE RIBOSOMAL PROTEIN.
DE F13F21.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC eumolliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
OC Arabidopsis.
OC [1]
RN SEQUENCE FROM N.A.
RR FEDERSPIEL N.A., PALM C.J., CONWAY A.B., CONN L., HANSEN N.F.,
RA ALTAFI H., ARAUJO R., HUIZAR I., ROWLEY D., BUEHLER E., DUNN P.,
RC GONZALEZ A., KREMENTSKAIA I., KIM C., LENZ C., LI J., LIU S.,
RA LURGS S., SCHWARTZ J., SHINN P., TORIUMI M., VYSOTSKAIA V.S.,
RRL WALKER M., YU G., ECKER J., THEOLOGIS A., DAVIS R.W.;
RRL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
EMBL; AC007504; AAD43160.1; -.
DSQ Ribosomal protein.
DSQ SEQUENCE 116 AA; 12560 MW; 08946EDD CRC32;

Query Match           29.0%; Score 70; DB 10; Length 116;
Best Local Similarity 21.9%; Pred. No. 1.31e-01;
Matches 7; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Ddb      32 RVVKTSKFMHDDDACNIGDRVKLDPSPRL 63
        ||||| : : || : : : : : : :
QY       1 QYKANSFIGITELGGCHGSEPCNIHRGKPF 32

RESULT      10
O024118     PRELIMINARY;          PRT;       102 AA.
AAC         O24118;
ACD         01-JAN-1998 (TReMBLrel. 05, Created)
DDT         01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DTI         01-NOV-1999 (TReMBLrel. 12, Last annotation update)
DE 40S RIBOSOMAL PROTEIN S17.

```

PETERS S.A., VAN STAVEREN M., DIRKSE W., STIEKEMA W., BANCROFT I., MEWES H.W., MAYER K.F.X., SCHUELLER C., BEVAN M.; Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

EU ARABIDOPSIS SEQUENCING PROJECT; Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.

EMBL; AL031018; CAA19819.1; PFAM; PF01554; DP0013; 1. Hypothetical protein.

KW Hypothetical protein.

SQ SEQUENCE 502 AA; 55232 MW; B6DF0F26 CRC32;

Query Match 28.6%; Score 69; DB 10; Length 502;
Best Local Similarity 42.9%; Pred. No. 2.12e-01;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 382 IIGLCELGNCPOTTLCGVLRG 402
: ||: ||| | : | : ||
9 FIGITELGGCHGSEPCNIHRG 29

RESULT 13 PRELIMINARY; PRT; 62 AA.

ID Q9X5H5;
AC Q9X5H5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DR EMBL; AL031018; CAA19819.1; PFAM; PF01554; DP0013; 1.
DE UREASE-ENHANCING FACTOR LPP.
OS LPP.
GN Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 43504;
RX MEDLINE; 99214098.
RA MCGEE D.J., MAY C.A., GARNER R.M., HIMPSE J.M., MOBLEY H.L.T.;
RT "Isolation of Helicobacter pylori genes that modulate urease activity".
RL J. Bacteriol. 181:2477-2484(1999).
RS EMBL; AF125214; AAD27692.1;
SQ SEQUENCE 62 AA; 6812 MW; FOBA2DES CRC32;

Query Match 28.2%; Score 68; DB 2; Length 62;
Best Local Similarity 58.8%; Pred. No. 3.40e-01;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 6 NSMFIGASLGGCASVE 22
||: ||| : ||| : |
6 NSKFIGITELGGCHGSE 22

RESULT 14 PRELIMINARY; PRT; 382 AA.

ID O32874;
AC O32874;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DR EMBL; AL031018; CAA19819.1; PFAM; PF01554; DP0013; 1.
DE HOMOSERINE O-ACETYLTRANSFERASE.
OS MCB1779.11.
GN Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA HAMILIN N., CHURCHER C.M.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

Search completed: Sun Apr 2 18:09:54 2000
Job time : 73 secs.

MPERCH_PP

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:07:52 2000; MasPar time 28.47 Seconds

Tabular output not generated. 33.572 Million cell updates/sec

Title: >US-09-362-731-5

Description: (1-32) from US09362731.pep

Perfect Score: 241

Sequence: 1 QYKANSKFIGITELGCGHGPCNIHRKPF 32

Scoring table: PAM 150

Gap 11

Searched: 82229 seqs, 29864866 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: swiss-prot38

1:swissprot

Statistics: Mean 31.137; Variance 37.361; scale 0.833

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	122	50.6	146	1	DER2_DERPT MITE ALLERGEN DER P 2	6.89e-15
2	122	50.6	146	1	DEF2_DERFA MITE ALLERGEN DER F 2	6.89e-15
3	104	43.2	1314	1	TETX_CLOTE TETANUS TOXIN PRECURSO	4.44e-10
4	71	29.5	713	1	DDX4_RAT DEAD BOX PROTEIN 4 (VA	3.47e-02
5	69	28.6	141	1	ALG2_TYRPU GROUP 2 ALLERGEN PREC	9.19e-02
6	69	28.6	358	1	MET2_HAEIN PUTATIVE HOMOSERINE O-	9.19e-02
7	67	27.8	141	1	LEP1_LEPDS MITE ALLERGEN LEP D 1	2.38e-01
8	66	27.4	149	1	EPI1_CANFA EPIDIDYMAL SECRETORY P	3.81e-01
9	65	27.0	151	1	EPI1_HUMAN EPIDIDYMAL SECRETORY P	6.06e-01
10	64	26.6	149	1	EPI1_BOVIN SECRETORY PROTEIN E1 P	9.59e-01
11	64	26.6	271	1	MURI_MYCTU PROBABLE GLUTAMATE RAC	9.59e-01
12	64	26.6	333	1	DOEB_XENLA DNA POLYMERASE BETA (E	9.59e-01
13	64	26.6	378	1	YD40_YEAST HYPOTHETICAL 42.3 KD P	9.59e-01
14	63	26.1	213	1	KAD_MYCCA ADENYLATE KINASE (EC 2	1.51e+00
15	63	26.1	261	1	EIA_ADE07 EARLY EIA 28 KD PROTEI	1.51e+00
16	63	26.1	899	1	V120_HSVSA CAPSID ASSEMBLY PROTEI	1.51e+00
17	62	25.7	284	1	RIBF_SYNY3 RIBOPOLYMER BIOSYNTHESI	2.36e+00
18	62	25.7	287	1	TRUB_AQUAE TRNA PSEUDOURIDINE SYN	2.36e+00
19	62	25.7	339	1	TCMO_STRGA TETRACENOMYCIN POLYKET	2.36e+00
20	62	25.7	398	1	RFE_MYCLE PUTATIVE UNDECAPEPTYL-	2.36e+00
21	62	25.7	404	1	RFE_MYCTU PUTATIVE UNDECAPEPTYL-	2.36e+00
22	62	25.7	608	1	RDPO_SCOEB PROBABLE REVERSE TRANS	2.36e+00
23	62	25.7	2156	1	ORP1_HUMAN OXYGEN-REGULATED PROTE	2.36e+00

Query Match 50.6%; Score 122; DB 1; Length 146;
Best Local Similarity 93.8%; Pred. No. 6.89e-15;

24	61	25.3	194	1	ACPD_HABIN ACYL CARRIER PROTEIN P	3.67e+00
25	61	25.3	547	1	NLTP_CHICK NONSPECIFIC LIPID-TRAN	3.67e+00
26	61	25.3	551	1	LEM2_RABIT E-SELECTIN PRECURSOR (3.67e+00
27	61	25.3	637	1	DDX4_MOUSE DEAD BOX PROTEIN 4 (VA	3.67e+00
28	61	25.3	859	1	OBP_HSVBC REPLICATION ORIGIN BIN	3.67e+00
29	61	25.3	3075	1	LMA1_HUMAN LAMININ ALPHA-1 CHAIN	3.67e+00
30	60	24.9	269	1	Y579_METJA LAMININ ALPHA-1 CHAIN	3.67e+00
31	60	24.9	453	1	UCR2_HUMAN UBIQUINOL-CYTOCHROME C	5.67e+00
32	60	24.9	568	1	GGT_PIG GAMMA-GLUTAMYLTRANSFER	5.67e+00
33	60	24.9	728	1	VIV_ORISA VIVIPAROUS PROTEIN HOM	5.67e+00
34	60	24.9	3084	1	LMA1_MOUSE LAMININ ALPHA-1 CHAIN	5.67e+00
35	59	24.5	125	1	LEEB_NPYOP LATE EXPRESSION FACTOR	8.71e+00
36	59	24.5	148	1	TRBH_RHISN PROBABLE CONJUGAL TRAN	8.71e+00
37	59	24.5	305	1	YHDM_ECOLI HYPOTHETICAL AMINO-ACI	8.71e+00
38	59	24.5	348	1	CTGF_MOUSE CONNECTIVE TISSUE GROW	8.71e+00
39	59	24.5	349	1	CTGF_PIG CONNECTIVE TISSUE GROW	8.71e+00
40	59	24.5	349	1	CTGF_HUMAN CONNECTIVE TISSUE GROW	8.71e+00
41	59	24.5	449	1	YKE6_YEAST HYPOTHETICAL 49.6 KD P	8.71e+00
42	59	24.5	537	1	MIPH_CHICK MYOSIN-BINDING PROTEIN	8.71e+00
43	59	24.5	635	1	YGC3_YEAST HYPOTHETICAL 70.6 KD P	8.71e+00
44	59	24.5	1227	1	PR16_HUMAN PRE-MRNA SPLICING FACT	8.71e+00
45	59	24.5	4393	1	PGBM_HUMAN BASEMENT MEMBRANE-SPEC	8.71e+00

ALIGNMENTS

RESULT 1
ID DER2_DERPT STANDARD; PRT; 146 AA.
AC P49278;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MITE ALLERGEN DER P 2 PRECURSOR (DER P II) (DPX).
GN DERP2.
OS Dermatophagoides pteronyssinus (House-dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
OC Dermatophagoides.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90256301.
RA CHUA K.I., DOYLE C.R., SIMPSON R.J., TURNER K.J., STEWART G.A.,
RA THOMAS W.R.;
RT "Isolation of cDNA coding for the major mite allergen Der p II by IgE
RT plaque immunoassay."
RL Int. Arch. Allergy Appl. Immunol. 91:118-123(1990).
RN [2]
RP PARTIAL SEQUENCE OF 18-57.
RX MEDLINE; 89278484.
RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
RA PLATTIS-MILLS T.A.;
RT "Antigenic and structural analysis of group II allergens (Der f II
RT and Der p II) from house dust mites (Dermatophagoides spp).";
RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE; 98409423.
RA MUELLER G.A., BENJAMIN D.C., RULE G.S.;
RT "Tertiary structure of the major house dust mite allergen Der p 2:
RT sequential and structural homologies.";
RL Biochemistry 37:12707-12714(1998).
CC 1- SIMILARITY: TO MITE ALLERGEN LEP D I.
DR PDB; 1ASV; 14-OCT-98.
KW Allergen; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 146 MITE ALLERGEN DER P 2.
FT DISULFID 25 136
FT DISULFID 38 44
FT DISULFID 90 95
SQ SEQUENCE 146 AA; 15999 MW; 09A45F2E CRC32;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 37 GCGSEPCIIHRGKPF 52
 11111111 11111111

Qy 17 GCGSEPCIIHRGKPF 32

RESULT 2

ID DEF2_DERFA STANDARD; PRT; 146 AA.
 AC Q00855; P39672; Q26359;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE MITE ALLERGEN DER F 2 PRECURSOR (DER F II).
 GN DERF2.
 OS Dermatophagoides farinae (House-dust mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcophagiformes; Astigmata; Pyroglyphoidea; Pyroglyphidae;
 OC Dermatophagoides.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE: 91291341.
 RA YUKI T., OKUMURA Y., ANDO T., YAMAKAWA H., SUKO M., HAIDA M.,
 RA OKUDAIRA H.;
 RT "Cloning and expression of cDNA coding for the major house dust mite
 RT allergen Der f II in Escherichia coli.";
 RL Agric. Biol. Chem. 55:1233-1238(1991).
 RN [2]
 RP SEQUENCE OF 4-146 FROM N.A.
 RX MEDLINE: 94256850.
 RA OKUHARA H.;
 RT "Molecular biology of mite antigens.";
 RL Alerugi 43:435-440(1994).
 RN [3]
 RP DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE: 93283958.
 RA NISHIYAMA C., YUKI T., TAKAI T., OKUMURA Y., OKUDAIRA H.;
 RT "Determination of three disulfide bonds in a major house dust mite
 RT allergen, Der f II.";
 RL Int. Arch. Allergy Immunol. 101:159-166(1993).
 RN [4]
 RP PARTIAL SEQUENCE OF 18-52.
 RX MEDLINE: 89278484.
 RA HEYMANN P.W., CHAPMAN M.D., AALBERSE R.C., FOX J.W.,
 RA PLATTS-MILLS T.A.;
 RT "Antigenic and structural analysis of group II allergens (Der f II
 RT and Der p II) from house dust mites (Dermatophagoides spp).";
 RL J. Allergy Clin. Immunol. 83:1055-1067(1989).
 RN [5]
 RP STRUCTURE BY NMR.
 RX MEDLINE: 98079068.
 RA ICHIKAWA S., HATANAKA H., YUKI T., IWAMOTO N., KOJIMA S.,
 RA NISHIYAMA C., OGURA K., OKUMURA Y., INAGAKI F.;
 RT "Solution structure of Der f 2, the major mite allergen for atopic
 RT diseases.";
 RL J. Biol. Chem. 273:356-360(1998).
 CC -!- MISCELLANEOUS: THE SEQUENCE SHOWN HERE IS FROM CLONE 2. THE
 CC N-TERMINAL SEQUENCE (AA 1-8) FROM CLONE 1 AND 11 ARE NOT YET
 CC KNOWN.
 CC -!- SIMILARITY: TO MITE ALLERGEN LEP D I.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10447; BAA01239.1;
 CC EMBL: D10448; BAA01240.1;
 CC EMBL: D10449; BAA01241.1;
 CC EMBL: S70378; AAB30829.1;

DR PIR: PS0417; PS0417.
 DR PDB: 1AHK; 08-APR-98.
 DR PDB: 1AHM; 08-APR-98.
 KW Allergen; Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 17
 FT CHAIN 18 146 MITE ALLERGEN DER F 2.
 FT DISULFID 25 136
 FT DISULFID 38 44
 FT DISULFID 90 95
 FT VARIANT 93 93 M -> V (IN CLONE 11).
 FT VARIANT 105 105 I -> A (IN CLONE 11).
 FT VARIANT 128 128 I -> V (IN CLONE 11).
 FT VARIANT 142 142 G -> A (IN CLONE 11).
 FT CONFLICT 5 8 ILCL -> GTMV (IN REF. 2).
 SQ SEQUENCE 146 AA; 15802 MW; 72623F23 CRC32;
 Query Match 50.6%; Score 122; DB 1; Length 146;
 Best Local Similarity 77.8%; Pred. No. 6.89e-15;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 35 VDCGSGSDPCIIHRGKPF 52
 : 11111111 11111111
 Qy 15 LGGCHGSEPCIIHRGKPF 32

RESULT 3

ID TETX_CLOTE STANDARD; PRT; 1314 AA.
 AC P04958;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
 OS Clostridium tetani.
 OG Plasmid
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87053814.
 RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
 RA WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;
 RT "Tetanus toxin: primary structure, expression in E. coli, and
 RT homology with botulinum toxins.";
 RL EMBO J. 5:2495-2502(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CN3911.
 RX MEDLINE: 87040747.
 RA FAIRWEATHER N.F., LYNNESS V.A.;
 RT "The complete nucleotide sequence of tetanus toxin.";
 RL Nucleic Acids Res. 14:7809-7812(1986).
 RN [3]
 RP SEQUENCE OF 742-1314 FROM N.A.
 RX MEDLINE: 86085672.
 RA FAIRWEATHER N.F., LYNNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
 RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
 RT fragment C in Escherichia coli.";
 RL J. Bacteriol. 165:21-27(1986).
 RN [4]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE: 90201034.
 RA KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;
 RT "Arrangement of disulfide bridges and positions of sulfhydryl groups
 RT in tetanus toxin.";
 RL Eur. J. Biochem. 188:39-45(1990).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE: 92037649.
 RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;
 RT "Limited proteolysis of tetanus toxin. Relation to activity and
 RT identification of cleavage sites.";
 RL Eur. J. Biochem. 202:41-51(1991).
 RN [6]

RP IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE; 93010948.
RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
RA MONTECUCO C.;
RT "Tetanus toxin is a zinc protein and its inhibition of
RT neurotransmitter release and protease activity depend on zinc.";
RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 93063293.
RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
RA DASGUPTA B.R., MONTECUCO C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin.";
RL Nature 359:832-835(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE; 97475217.
RA UNLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
RA SAX M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin.";
RL Nat. Struct. Biol. 4:788-792(1997).
CC -!- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
CC BOND OF SYNAPTOSOMAL-2.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
CC SYNAPTOSOMAL-2.
CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -!- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO
CC GANGLIOSIDE RECEPTORS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04436; CAA28033.1; -
DR EMBL; M12739; AAA23282.1; -
DR EMBL; X06214; CAA29564.1; -
DR PIR; A25689; BTCLTN.
DR PDB; 1AF9; 29-APR-98.
DR PDB; 1ABD; 14-OCT-98.
DR PROSITE; PS00142; ZINC-PROTEASE; 1.
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;
KW 3D-structure.
FT INIT_MET 0
FT CHAIN 1 456
FT CHAIN 457 1314
FT ACT_SITE 232 232
FT ACT_SITE 233 233
FT ACT_SITE 236 236
FT METAL 226 246
FT TRANSMEM 669 689
FT TRANSMEM 669 689
FT DISULFID 438 466
FT DISULFID 1076 1092
FT SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;
Query Match 43.28; Score 104; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 4.44e-10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 829 QYIKANSKFIGITEL 843
QY 1 QYIKANSKFIGITEL 15
RESULT 4
ID DDX4_RAT STANDARD; PRT; 713 AA.
AC Q64060;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DEAD BOX PROTEIN 4 (VASA HOMOLOG) (RVLG).
GN DDX4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-TESTIS;
RX MEDLINE; 95160706.
RA KOMIYA T., TANIGAWA Y.;
RT "Cloning of a gene of the DEAD box protein family which is
RT specifically expressed in germ cells in rats.";
RL Biochem. Biophys. Res. Commun. 207:405-410(1995).
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES. HIGHEST TO
CC DROSOPHILA VASA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S75275; AAB33364.1; -
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
DR PFAM; PF00270; DEAD; 1.
DR PFAM; PF00271; Helicase_C; 1.
KW ATP-binding; Helicase; RNA-binding.
FT NP_BIND 317 324 ATP (POTENTIAL).
FT SITE 431 434 DEAD BOX.
SQ SEQUENCE 713 AA; 77955 MW; 34CEB6D3 CRC32;
Query Match 29.5%; Score 71; DB 1; Length 713;
Best Local Similarity 24.2%; Pred. No. 3.47e-02;
Matches 8; Conservative 13; Mismatches 11; Indels 1; Gaps 1;
Db 479 EFLKSNLYFVAVGVGGACRDVQSQILOVGPVF 511
QY 1 QYIKANSKFIGITELGG-CHGSEPCNIHRKPF 32
RESULT 5
ID ALG2_TYRPU STANDARD; PRT; 141 AA.
AC O02380;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GROUP 2 ALLERGEN PRECURSOR.
OS Tyrophagus putrescentiae (Dust mite).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcoptiformes; Astigmata; Acaroidea; Acaridae;
OC Tyrophagus.
RN [1]
RP SEQUENCE FROM N.A.
RA ERIKSSON T., JOHANSSON E., WHITLEY P., SCHMIDT M., ELSAYED S.,
RA VAN HAGE-HAMSTEN M.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE E1 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Y12690; CAA73221.1;
 DR HSP; Q00855; IAHM.
 KW Allergen; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 141
 FT DISULFID 23 132
 FT DISULFID 36 41
 FT DISULFID 87 92
 FT CARBOHYD 103 103
 FT SEQUENCE 141 AA; 14851 MW; 1635DCE8 CRC32;

Query Match 28.6%; Score 69; DB 1; Length 141;
 Best Local Similarity 40.9%; Pred. No. 9.19e-02;
 Matches 9; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

DB 28 IASVAVDCEG-DLCVTHKSKP 48
 QY 10 IGITELGGCHGSEPCNIHRGKP 31

RESULT 6
 ID MET2_HAEIN STANDARD; PRT; 358 AA.
 AC P45131;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE PUTATIVE HOMOSERINE O-ACETYLTRANSFERASE (EC 2.3.1.31) (HOMOSERINE O-
 DE TRANS-ACETYLASE).
 GN H11263.

OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., TOMB J.F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZGUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GHEHM C.L., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RL Science 269:496-512(1995).

CC -1- CATALYTIC ACTIVITY: ACETYL-COA + L-HOMOSERINE - COA + O-ACETYL-L-
 CC HOMOSERINE.
 CC -1- SIMILARITY: STRONG, TO FUNGAL HOMOSERINE O-ACETYLTRANSFERASE
 CC (MET2).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; U32806; AAC22916.1;
 DR TIGR; H11263;
 DR PFAM; PF00561; abhydrolase: 1.

KW Hypothetical protein; Methionine biosynthesis; Transferase;
 KW Acyltransferase.
 SQ SEQUENCE 358 AA; 39990 MW; 51EEF250 CRC32;

Query Match 28.6%; Score 69; DB 1; Length 358;
 Best Local Similarity 44.4%; Pred. No. 9.19e-02;
 Matches 12; Conservative 6; Mismatches 6; Indels 3; Gaps 3;

DB 82 FISSNVLGGKGTGTPSSINPQTGKPY 108
 QY 9 FIGITELGGCHGS-EPCNIH-R-GKPF 32

RESULT 7
 ID LEPI_LEPDS STANDARD; PRT; 141 AA.
 AC P80384;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MITE ALLERGEN LEP D 1 PRECURSOR (LEP D I.01 AND LEP D I.02).
 OS Lepidoglyphus destructor (Storage mite).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 OC Acariformes; Sarcotiformes; Astigmata; Glycyphagoidea; Glycyphagidae;
 OC Lepidoglyphus.

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 95377437.
 RA SCHMIDT M., VAN DER PLOEG I., OLSSON S., VAN HAGE HAMSTEN M.;
 RT "CDNA analysis of the mite allergen Lep d 1 identifies two different
 RT isocallergens and variants.";
 RL FEBS Lett. 370:11-14(1995).

RN [2]
 RP SEQUENCE OF 44-141 FROM N.A., AND SEQUENCE OF 17-140.
 RX MEDLINE; 95010146.
 RA VARELA J., VENTAS P., CARREIRA J., BARBAS J.A., GIMENEZ-GALLEGO G.,
 RA POLO F.;
 RT "Primary structure of Lep d I, the main Lepidoglyphus destructor
 RT allergen.";
 RL Eur. J. Biochem. 225:93-98(1994).

RN [3]
 RP PARTIAL SEQUENCE OF 17-45.
 RA MUTHIAH R., MILLER M., KAGEN S.;
 RT "Barn allergy: Isolation and characterization of the major allergens
 RT of storage mites: L. destructor.";
 RL J. Allergy Clin. Immunol. 87:326-326(1991).

RN [4]
 RP SEQUENCE OF 17-34.
 RX MEDLINE; 92382323.
 RA VAN HAGE-HAMSTEN M., BERGMAN T., JOHANSSON E., PERSSON B.,
 RA JOERNVALL H., HAERFEST B., JOHANSSON S.G.O.;
 RT "N-terminal amino acid sequence of principal allergen of storage mite
 RT Lepidoglyphus destructor.";
 RL Lancet 340:614-614(1992).

CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: TO MITE ALLERGEN DER F II.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X89014; CAA61419.1;
 DR EMBL; X83875; CAA58755.1;
 DR EMBL; X83876; CAA58756.1;
 DR EMBL; X81399; CAA57160.1;
 DR HSP; Q00855; IAHM.
 KW Allergen; Polymorphism; Repeat; Signal.
 FT SIGNAL 1 16
 FT CHAIN 17 141
 FT DOMAIN 64 73
 FT MITE ALLERGEN LEP D I.01.
 FT 3 X 2 AA REPEATS OF K-V.

FT REPEAT 64 65
 FT REPEAT 68 69
 FT REPEAT 72 73
 FT DISULFID 24 133
 FT DISULFID 37 42
 FT DISULFID 88 93
 FT VARIANT 35 35
 FT VARIANT 48 48
 FT VARIANT 53 53
 FT VARIANT 63 63
 FT VARIANT 90 91
 FT VARIANT 95 95
 FT VARIANT 104 104
 FT VARIANT 106 107
 FT VARIANT 116 116
 FT VARIANT 125 125
 FT VARIANT 136 136
 FT CONFLICT 26 26
 FT CONFLICT 30 30
 SQ SEQUENCE 141 AA; 14773 MW; 286421C CRC32;

Query Match 27.8%; Score 67; DB 1; Length 141;
 Best Local Similarity 50.0%; Pred. No. 2.38e-01;
 Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

Db 33 DITGCGG-DTCVIHRG 47
 : : | | : | | | |
 QY 14 ELGGCHGSEPCNIHRG 29

RESULT 8
 ID EPI_CANFA STANDARD; PRT: 149 AA.
 AC Q28895;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE EPIDIDYMAL SECRETORY PROTEIN EI PRECURSOR (CEI).
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=EPIDIDYMIS;
 RX MEDLINE; 95263175.
 RA ELLERBROCK K., PERA I., HARTUNG S., IVELL R.;
 RT "Gene expression in the dog epididymis: a model for human epididymal function";
 RL Int. J. Androl. 17:314-323(1994).
 CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: EPIDIDYMIS.
 CC -!- SIMILARITY: BELONGS TO THE EI FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S77411; AAB34263.1; -
 DR SIGNAL
 FT CHAIN 1 21 POTENTIAL.
 FT DISULFID 22 149 EPIDIDYMAL SECRETORY PROTEIN EI.
 FT DISULFID 27 140 BY SIMILARITY.
 FT DISULFID 42 47 BY SIMILARITY.
 FT DISULFID 93 99 BY SIMILARITY.
 FT CARBOHYD 58 58 POTENTIAL.
 SQ SEQUENCE 149 AA; 16056 MW; 2908708D CRC32;

Query Match 27.4%; Score 66; DB 1; Length 149;
 Best Local Similarity 33.3%; Pred. No. 3.81e-01;
 Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 87 IPEADGCKSGINCPIQDKTY 107
 : : | | : | | : | :
 QY 12 ITELGGCHGSEPCNIHRGKPF 32

RESULT 9
 ID EPI_HUMAN STANDARD; PRT: 151 AA.
 AC Q15668; Q29413;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE EPIDIDYMAL SECRETORY PROTEIN EI PRECURSOR (EPI-1) (HE1) (EPIDIDYMAL SECRETORY PROTEIN 14.6) (ESPI4.6).
 OS Homo sapiens (Human), Pan troglodytes (Chimpanzee), and Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=HUMAN; TISSUE=EPIDIDYMIS;
 RX MEDLINE; 93119659.
 RA KRULL N., IVELL R., OSTERHOFF C., KIRCHHOFF C.;
 RT "Region-specific variation of gene expression in the human epididymis as revealed by in situ hybridization with tissue-specific cDNAs";
 RL Mol. Reprod. Dev. 34:16-24(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=P. TROGLODYTES; TISSUE=EPIDIDYMIS;
 RX FROHLICH O., YOUNG L.G., MURPHY T.C.;
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M. FASCICULARIS; TISSUE=EPIDIDYMIS;
 RX MEDLINE; 95180740.
 RA PERRY A.C.F., JONES R., HALL L.;
 RT "The monkey ESPI4.6 mRNA, a novel transcript expressed at high levels in the epididymis";
 RL Gene 153:291-292(1995).
 CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: EPIDIDYMIS.
 CC -!- SIMILARITY: BELONGS TO THE EI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X67698; CAA47928.1; -
 DR EMBL; A18921; CAA01431.1; -
 DR EMBL; U25748; AAA67077.1; -
 DR EMBL; X78134; CAA55013.1; -
 KW Signal.
 FT CHAIN 1 19 POTENTIAL.
 FT DISULFID 20 151 EPIDIDYMAL SECRETORY PROTEIN EI.
 FT DISULFID 27 140 BY SIMILARITY.
 FT DISULFID 42 47 BY SIMILARITY.
 FT DISULFID 93 99 BY SIMILARITY.
 FT CARBOHYD 58 58 POTENTIAL.
 FT CARBOHYD 135 135 POTENTIAL.
 SQ SEQUENCE 151 AA; 16570 MW; 2EC86F5 CRC32;

Query Match 27.0%; Score 65; DB 1; Length 151;
 Best Local Similarity 33.3%; Pred. No. 6.06e-01;
 Matches 7; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Db 87 IPEPDGCKSGINCPIQDKTY 107
 : : | | : | | : | :
 QY 12 ITELGGCHGSEPCNIHRGKPF 32

RESULT 10
ID EPL_BOVIN STANDARD; PRT; 149 AA.
AC F79345;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SECRETORY PROTEIN EI PRECURSOR (EPV20).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RC TISSUE-MAMMARY GLAND;
RX MEDLINE; 97182631.
RA LARSEN L.B., RAVN P., BOISEN A., BERGLUND L., PETERSEN T.E.;
RT "Primary structure of EPV20, a secretory glycoprotein containing a
previously uncharacterized type of domain."
RL Eur. J. Biochem. 243:437-441(1997).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY, SPLEEN, LIVER AND MAMMARY
GLAND, BUT NOT IN TESTIS.
CC -1- SIMILARITY: BELONGS TO THE EI FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; X85799; CAA59794.1;
KW Signal.
FT SIGNAL 1 19
FT CHAIN 20 149 SECRETORY PROTEIN EI.
FT DISULFID 27 140
FT DISULFID 42 47
FT DISULFID 93 99
FT CARBOHYD 58 58 POTENTIAL.
SQ SEQUENCE 149 AA; 16640 MW; AFF60F1 CRC32;
Query Match 26.6%; Score 64; DB 1; Length 149;
Best Local Similarity 31.6%; Pred. No. 9.59e-01;
Matches 6; Conservative 9; Mismatches 3; Indels 1; Gaps 1;
Db 38 NVSPCP-TQCKLHGQSY 55
QY 14 ELGGCHGSEPCNHRGKPF 32

RESULT 11
ID MURI_MYCTU STANDARD; PRT; 271 AA.
AC Q10626;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 39, Last annotation update)
DE PROBABLE GLUTAMATE RACEMASE (EC 5.1.1.3).
GN MURI OR RV1338 OR MTCY130.23 OR MTCY02B10.02.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE; 98295987.
RA COLE S.T., BROCH R., PARKHILL J., GARNIER T., CHURCHER C., HARRIS D.,
RA GORDON S.V., EIGLMEIER K., GAS S., BARRY C.E. III, TEKAIA F.,
RA BADCOCK K., BASHAM D., BROWN D., CHILLINGWORTH T., CONNOR R.,
RA DAVIES R., DEVLIN K., FELTWELL T., GENTLES S., HAWLIN N., HOLROYD S.,
RA HORNSBY T., JAGELS K., KROGH A., MCLEAN J., MOULE S., MURPHY L.,
RA OLIVER S., OSBORNE J., QUAIL M.A., RAJANDREAM M.A., ROGERS J.,

RA RUTTER S., SEEGER K., SKELTON S., SQUARES S., SQUARES R., SULSTON J.E.,
RA TAYLOR K., WHITEHEAD S., BARRELL B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- FUNCTION: PROVIDES THE (R)-GLUTAMIC ACID REQUIRED FOR CELL
WALL BIOSYNTHESIS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE - D-GLUTAMATE.
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ASPARTATE/GLUTAMATE RACEMASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; Z73902; CAA98102.1;
DR PROSITE; PS00923; ASP_GLU_RACEMASE_1; 1.
DR PROSITE; PS00924; ASP_GLU_RACEMASE_2; 1.
DR PFAM; PF01177; ASP_Glu_race; 1.
KW Peptidoglycan synthesis; Cell wall; Isomerase.
SQ SEQUENCE 271 AA; 28643 MW; 2219C126 CRC32;
Query Match 26.6%; Score 64; DB 1; Length 271;
Best Local Similarity 34.6%; Pred. No. 9.59e-01;
Matches 9; Conservative 5; Mismatches 11; Indels 1; Gaps 1;
Db 247 FTKLAARFLG-PVLGGVQVPVHPSRIH 271
QY 2 YIRANSFTEIGTGCGHSEPCNIH 27
RESULT 12
ID DPOB_XENLA STANDARD; PRT; 333 AA.
AC Q57383;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA POLYMERASE BETA (EC 2.7.7.7).
GN POLB.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
OC Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98151235.
RA REICHENBERGER S., PFEIFFER P.;
RT "Cloning, purification and characterization of DNA polymerase beta
from Xenopus laevis -- studies on its potential role in DNA-end
joining."
RL Eur. J. Biochem. 251:81-90(1998).
CC -1- FUNCTION: REPAIR POLYMERASE. CONDUCTS "GAP-FILLING" DNA SYNTHESIS
IN A STEPWISE DISTRIBUTIVE FASHION RATHER THAN IN A PROCESSIVE
FASHION AS FOR OTHER DNA POLYMERASES.
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
N PYROPHOSPHATE + DNA(N).
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

```
CC EMBL; Y15732; CAA75741.1; -
DR PROSITE; PS00522; DNA_POLYMERASE_X; 1.
DR PFAM; PF00966; DNA_PolymeraseX; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 182 182 INVOLVED IN PRIMER BINDING
FT ACT_SITE 189 189 (BY SIMILARITY).
FT ACT_SITE 189 189 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 (BY SIMILARITY).
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 (BY SIMILARITY).
SQ SEQUENCE 333 AA; 38162 MW; 6948CECC CRC32;

Query Match 26.6%; Score 64; DB 1; Length 333;
Best Local Similarity 26.9%; Pred. No. 9.59e-01;
Matches 7; Conservative 10; Mismatches 8; Indels 1; Gaps 1;

Db 228 VKGDKFMGVCLP-CESDQDPYRR 252
QY 3 IKANSKFIGITELGGCHGSEPCNIHR 28

RESULT 13
ID YD40.YEAST STANDARD; PRT; 378 AA.
AC Q04179;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 42.3 KD PROTEIN IN YTA2-DIT1 INTERGENIC REGION.
GN YD400W OR D9509.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
RA BERNO A., CARPENTER J., CHEN E., CHERRY J.M., CHUNG E., DUNCAN M.,
RA HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAWATH A., OEFNER P., OH C., PETEL F.X.,
RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGREN T., SHROFF N.,
RA WINANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
RA Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RL
CC -1- SIMILARITY: BELONGS TO THE INH FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X06414; CAA29724.1; -
DR PIR; S02851; KIYMC.
DR HSSP; P27142; 12IO.
DR PROSITE; PS00113; ADENYLATE KINASE; 1.
DR PFAM; PF00406; adenylatekinase; 1.
KW Transferase; Kinase; ATP-binding.
FT NP_BIND 7 15 ATP (BY SIMILARITY).
SQ SEQUENCE 213 AA; 24616 MW; 8D5C4B03 CRC32;

Query Match 26.1%; Score 63; DB 1; Length 213;
Best Local Similarity 66.7%; Pred. No. 1.51e-00;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 182 DYFKTNSKFIEI 193
QY 1 QVIKANSKFIGI 12

RESULT 15
ID E1A.ADR07 STANDARD; PRT; 261 AA.
AC P03256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE EARLY E1A 28 KD PROTEIN [CONTAINS: EARLY E1A 24 KD PROTEIN].
OS Human adenovirus type 7.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GONEN;
RX MEDLINE; 81237792.
RA DIJKEMA R., DEKKER B.M.M., VAN ORMONDT H., DE WAARD A., MAAT J.,
RA BOYER H.W.;
RT "Gene organization of the transforming region of weakly oncogenic
RT adenovirus type 7: the E1a region.";
RL Gene 12:287-299(1980).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-GRIDER;
RA YOSHIDA K., FUJINAGA K.;
RT "The nucleotide sequence of the transforming HindIII-I.J fragment of
RT adenovirus type 7 DNA.";
RL Tumor Res. 19:39-47(1984).
```

 W A P E R E H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sun Apr 2 18:06:45 2000; Maspar time 44.77 Seconds
 33.718 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-362-731-5
 Description: (1-32) from US09362731.ppe
 Perfect Score: 241
 Sequence: 1 QV1KANSKFTIGITELGCHGSEPCNIHRGPF 32

Scoring table: PAM 150
 Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: p1r62
 1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 30.417; Variance 42.186; scale 0.721

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.	
1	122	50.6	129	2	A61501	allergen Der f II - h	1.40e-12
2	122	50.6	129	2	JU0394	allergen Der f II (PF)	1.40e-12
3	122	50.6	138	2	A61241	allergen Der f II pre	1.40e-12
4	122	50.6	138	2	B61241	allergen Der f II pre	1.40e-12
5	122	50.6	146	2	A60381	allergen Der p II pre	1.40e-12
6	104	43.2	1315	1	BTCLTN	tentoxylisin (EC 3.4.	2.26e-08
7	75	31.1	533	2	A47143	retinal pigment micro	3.39e-02
8	72	29.9	173	2	T00970	hypothetical protein	1.28e-01
9	71	28.5	713	2	JC2534	RVG protein - rat	1.99e-01
10	71	28.5	867	2	T14777	hypothetical protein	1.99e-01
11	69	28.6	102	2	T15978	ribosomal protein S17	4.71e-01
12	69	28.6	141	2	S66499	allergen Lep d 1.02 p	4.71e-01
13	69	28.6	358	2	D64113	homoserine O-acetyltr	4.71e-01
14	69	28.6	502	2	T05135	hypothetical protein	4.71e-01
15	68	28.2	1036	2	B63368	hypothetical protein	7.20e-01
16	67	27.8	141	2	S65500	allergen Lep d 1.01 p	1.10e+00
17	66	27.4	149	2	I63229	CEL - dog	1.10e+00
18	66	27.4	379	2	D70846	probable metaA protein	1.66e+00
19	65	27.0	151	2	I53929	epididymal secretory	2.51e+00
20	65	27.0	151	2	I33365	epididymal secretory	2.51e+00
21	64	26.6	38	2	G64583	hypothetical protein	3.78e+00
22	64	26.6	159	2	A72660	hypothetical protein	3.78e+00
23	64	26.6	271	2	F70771	probable glutamaterac	3.78e+00

24	26.6	378	2	S69683	hypothetical protein	3.78e+00
25	26.1	213	1	KIYMC	adenylate kinase (EC	5.66e+00
26	26.1	261	1	WMAD87	early EIA 28K protein	5.66e+00
27	26.1	899	2	G36812	hypothetical protein	5.66e+00
28	26.1	183	2	B72225	conserved hypothetical	8.42e+00
29	25.7	219	1	A69505	uridylyate kinase (pyr	8.42e+00
30	25.7	284	2	F77138	hypothetical protein	8.42e+00
31	25.7	287	2	F70361	tRNA-pseudouridine sy	8.42e+00
32	25.7	339	2	C42276	O-methyltransferase -	8.42e+00
33	25.7	398	2	T09982	rfe protein - Mycobac	8.42e+00
34	25.7	404	2	B70774	probable rfe protein	8.42e+00
35	25.7	598	2	F69792	hypothetical protein	8.42e+00
36	25.7	608	2	S05341	probable reverse tran	8.42e+00
37	25.3	118	2	T11244	ribosomal protein S11	1.25e+01
38	25.3	156	2	S74049	hypothetical protein	1.25e+01
39	25.3	297	2	H75132	hypothetical protein	1.25e+01
40	25.3	361	2	T12571	cinnamyl-alcohol dehy	1.25e+01
41	25.3	433	2	B72025	flagellum-specific AT	1.25e+01
42	25.3	547	2	S34744	sterol carrier protei	1.25e+01
43	25.3	551	2	I46708	endothelial leukocyte	1.25e+01
44	25.3	551	2	I46709	endothelial leukocyte	1.25e+01
45	25.3	3075	2	S14458	laminin alpha-1 chain	1.25e+01

ALIGNMENTS

RESULT 1
 ENTRY A61501 #type fragment
 TITLE allergen Der f II - house-dust mite (Dermatophagoides
 ORGANISM farinae) (fragment)
 DATE 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change
 13-Sep-1998

ACCESSIONS A61501
 REFERENCE A61501
 #authors Trudinger, M.; Chua, K.Y.; Thomas, W.R.
 #journal Clin. Exp. Allergy (1991) 21:33-37
 #title CDNA encoding the major mite allergen Der f II.
 #cross-references MUID:91215495
 #accession A61501

##status preliminary; not compared with conceptual translation
 ##molecule_type mRNA
 ##residues 1-129 ##label TRU
 CLASSIFICATION #superfamily allergen Der p II
 SUMMARY #length 129 #checksum 476

Query Match 50.6%; Score 122; DB 2; Length 129;
 Best Local Similarity 77.8%; Pred. No. 1.40e-12;
 Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db	18	VDCGSGSDPCIIHRGPF	35
QY	15	LGCGHSEPCNIHRGPF	32

RESULT 2
 ENTRY JU0394 #type complete
 TITLE allergen Der f II (pFL2) - house-dust mite (Dermatophagoides
 ORGANISM farinae)
 DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change
 17-Mar-1999

ACCESSIONS JU0394
 REFERENCE PS0417
 #authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
 Haide, M.; Okudaira, H.
 #journal Agric. Biol. Chem. (1991) 55:1233-1238
 #title Cloning and expression of cDNA coding for the major house
 dust mite allergen Der f II in Escherichia coli.
 #cross-references MUID:91291341
 #accession JU0394
 ##molecule_type mRNA
 ##residues 1-129 ##label YU0


```

CLASSIFICATION #superfamily allergen Der p II
SUMMARY #length 129 #molecular-weight 14076 #checksum 9516

Query Match 50.6%; Score 122; DB 2; Length 129;
Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 18 VDCGSGSDPCIHRGKPF 35
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||: || |||||

RESULT 3
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS A61241
REFERENCE A61241; PS0417
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession A61241
#molecule_type mRNA
#residues 1-138 #label YUU
#note part of this sequence, including the amino end of the
mature protein, was confirmed by protein sequencing
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
10-138 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status experimental #label
MAT
SUMMARY #length 138 #checksum 2894

Query Match 50.6%; Score 122; DB 2; Length 138;
Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 27 VDCGSGSDPCIHRGKPF 44
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||: || |||||

RESULT 4
ENTRY A61241 #type fragment
TITLE allergen Der f II precursor - house-dust mite
(Dermatophagoides farinae) (fragment)
ORGANISM #formal_name Dermatophagoides farinae
DATE 12-May-1994 #sequence_revision 27-Jun-1994 #text_change
13-Sep-1998
ACCESSIONS B61241; J00395
REFERENCE A61241
#authors Yuuki, T.; Okumura, Y.; Ando, T.; Yamakawa, H.; Suko, M.;
Haida, M.; Dohi, M.; Okudaira, H.
#journal Int. Arch. Allergy Appl. Immunol. (1991) 94:354-356
#title Synthesis of biologically active recombinant Der f II.
#cross-references MUID:92040281
#accession B61241
#molecule_type mRNA
#residues 1-138 #label YUU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-9
10-138 #domain signal sequence (fragment) #status predicted
#label SIG\
#product allergen Der f II #status predicted #label MAT
SUMMARY #length 138 #checksum 2280

Query Match 50.6%; Score 122; DB 2; Length 138;

```

```

Best Local Similarity 77.8%; Pred. No. 1.40e-12;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 27 VDCGSGSDPCIHRGKPF 44
QY 15 LGCHGSEPCNIHRGKPF 32
: |||||: || |||||

RESULT 5
ENTRY A60381 #type complete
TITLE allergen Der p II precursor - house-dust mite
(Dermatophagoides pteronyssinus)
ORGANISM #formal_name Dermatophagoides pteronyssinus
DATE 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
13-Sep-1998
ACCESSIONS A60381
REFERENCE A60381
#authors Chua, K.Y.; Doyle, C.R.; Simpson, R.J.; Turner, K.J.;
Stewart, G.A.; Thomas, W.R.
#journal Int. Arch. Allergy Appl. Immunol. (1990) 91:118-123
#title Isolation of cDNA coding for the major mite allergen Der p II
by IgE plaque immunoassay.
#cross-references MUID:90256301
#accession A60381
#status not compared with conceptual translation
#molecule_type mRNA
#residues 1-146 #label CHU
CLASSIFICATION #superfamily allergen Der p II
FEATURE 1-17
18-146 #domain signal sequence #status predicted #label SIG\
#product allergen Der p II #status predicted #label MAT
SUMMARY #length 146 #molecular-weight 15999 #checksum 25

Query Match 50.6%; Score 122; DB 2; Length 146;
Best Local Similarity 93.8%; Pred. No. 1.40e-12;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 37 GCHGSEPCNIHRGKPF 52
QY 17 GCHGSEPCNIHRGKPF 32
: |||||: || |||||

RESULT 6
ENTRY BTCLTN #type complete
TITLE tentoxylisin (EC 3.4.24.68) precursor - Clostridium tetani
#formal_name Clostridium tetani
ORGANISM #formal_name Clostridium tetani
DATE 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
18-Jun-1999
ACCESSIONS A25689; A25757; A25194; B25194; A60759; S69348; S09364
REFERENCE A25689
#authors Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels,
J.; Weiler, U.; Hudel, M.; Habermann, E.; Niemann, H.
#journal EMBO J. (1986) 5:2495-2502
#title Tetanus toxin: primary structure, expression in E. coli, and
homology with botulinum toxins.
#cross-references MUID:87053814
#accession A25689
#molecule_type DNA
#residues 1-1315 #label EIS
#cross-references GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770
REFERENCE A25757
#authors Fairweather, N.F.; Lyness, V.A.
#journal Nucleic Acids Res. (1986) 14:7809-7812
#title The complete nucleotide sequence of tetanus toxin.
#cross-references MUID:87040747
#accession A25757
#molecule_type DNA
#residues 1-1315 #label FAI
#cross-references GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774
#experimental_source strain CN3911
REFERENCE A25194
#authors Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.;

```

Thomson, R.O.
#journal J. Bacteriol. (1986) 165:21-27
#title Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in *Escherichia coli*.
#cross-references MUID:86085672
#accession A25194
#molecule_type DNA
#residues 743-1315 #label FA2
#cross-references GB:M12739; NID:G144920; PIDN:AAA23282.1; PID:G144921
#accession B25194
#molecule_type protein
#residues 865-894 #label FA3
REFERENCE
#authors Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.
#journal Infect. Immun. (1989) 57:3588-3593
#title Isolation, purification, and characterization of fragment B, the NH-2-terminal half of the heavy chain of tetanus toxin.
#cross-references MUID:90035436
#accession A60759
#molecule_type protein
#residues 461-475 #label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell epitopes.
#cross-references MUID:89093918
#contents annotation; epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; Dasgupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
#molecule_type protein
#residues 2-31 #label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form light (fragment A) and heavy (fragment B.C) chains that are covalently linked by an interchain disulfide bond (the individual chains are not toxic when separated). The amino end of the heavy chain (fragment B) can be separated from the carboxyl end (fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds to gangliosides and may target the toxin to the motor end plate. Fragment A is a zinc-dependent endopeptidase.
COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is internalized, and moves by retrograde transport up the axon into the spinal cord where it can move between postsynaptic and presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of synaptobrevin (vesicle-associated membrane protein 2).
FUNCTION
#description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
FEATURE
2-457 #product tentoxylisin light chain (fragment A) #status predicted #label TTL
461-1315 #product tentoxylisin heavy chain (fragment B.C) #status

experimental #label TTH
#domain channel forming (fragment B) #status predicted
#label TXB
#domain ganglioside binding (fragment C) #status predicted
#binding_site zinc (His) #status predicted
#active_site Glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 43.2%; Score 104; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 2.26e-08;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 830 QYIKANSKFITIGTEL 844
QY 1 QYIKANSKFITIGTEL 15
|||||
RESULT 7
ENTRY A47143 #type complete
TITLE retinal pigment microsomal protein RPE65, epithelium-specific - bovine
ALTERNATE_NAMES membrane receptor p63; retinol-binding protein receptor
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change 10-Sep-1997
ACCESSIONS A47143; A48017; S28503
REFERENCE A47143
#authors Hamel, C.P.; Tsilou, E.; Pfeiffer, B.A.; Hooks, J.J.; Detrick, B.; Redmond, T.M.
#journal J. Biol. Chem. (1993) 268:15751-15757
#title Molecular cloning and expression of RPE65, a novel retinal pigment epithelium-specific microsomal protein that is post-transcriptionally regulated in vitro.
#cross-references MUID:93340181
#accession A47143 preliminary
#status #molecule_type mRNA
#residues 1-533 #label HAM
#cross-references GB:L11356; NID:G163656; PID:G163657
REFERENCE A48017
#authors Baviak, C.O.; Levy, F.; Hellman, U.; Wernstedt, C.; Eriksson, U.
#journal J. Biol. Chem. (1993) 268:20540-20546
#title The retinal pigment epithelial membrane receptor for plasma retinol-binding protein. Isolation and cDNA cloning of the 63-kDa protein.
#cross-references MUID:93388633
#accession A48017 preliminary
#status #molecule_type mRNA
#residues 1-340, 'L', 342-533 #label BA2
#cross-references EMBL:X66277; NID:G563; PID:G564
KEYWORDS membrane protein
SUMMARY #length 533 #molecular-weight 60944 #checksum 3885
Query Match 31.1%; Score 75; DB 2; Length 533;
Best Local Similarity 44.0%; Pred. No. 3.39e-02;
Matches 11; Conservative 7; Mismatches 6; Indels 1; Gaps 1;
Db 89 YVRAMTEKRIVITEFGTCAPDPCK 113
Y 2 YIKANS-KFTIGTELGGCHGSEPCN 25
|||||
RESULT 8
ENTRY T00970 #type complete
TITLE hypothetical protein T9J22.4 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 20-Sep-1999
ACCESSIONS T00970

```

20-Sep-1999
T14777
Z18183
Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.;
Wiemann, S.
Submitted to the Protein Sequence Database, August 1999
#submission
#accession T14777
#status preliminary
#molecule_type mRNA
#residues 1-867 #label POU
#cross-references EMBL:AL10249
#experimental_source adult testis; clone DKF2p434N061
GENETICS
#note
#summary DKF2p434N061.1
#length 867 #checksum 3074

Query Match 29.5% Score 71; DB 2; Length 867;
Best Local Similarity 44.0%; Pred. No. 1.99e-01;
Matches 11; Conservative 6; Mismatches 7; Indels 1; Gaps 1;

Db 432 IESDKVITISVTGLPGCHASELCAF 456
QY 3 IKANSKFIGI-TELGCGHSEPCNI 26
::: ||: || |||||:::

RESULT 11
ENTRY T16978 #type complete
TITLE ribosomal protein S17 - curled-leaved tobacco
ORGANISM #formal_name Nicotiana glumbaginifolia #common_name
curled-leaved tobacco
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
T16978
Z18621
Borisjuk, N.V.
#submission submitted to the EMBL Data Library, October 1996
#accession T16978
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-102 #label BOR
#cross-references EMBL:Y08858; NID:el008001; PID:e275056
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 102 #molecular-weight 11602 #checksum 6192

Query Match 28.6% Score 69; DB 2; Length 102;
Best Local Similarity 21.9%; Pred. No. 4.71e-01;
Matches 7; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

Db 32 RYVKRTSKFMAHDQDNCNIGDRVKLDHSRPL 63
QY 1 QYIRANSKFIGITELGGCHGSEPCNIHRGKPF 32
::: ||||: : ||: : ::::
1 QYIRANSKFIGITELGGCHGSEPCNIHRGKPF 32

RESULT 12
ENTRY S66499 #type complete
TITLE allergen Lep d 1.02 precursor - Lepidoglyphus destructor
ORGANISM #formal_name Lepidoglyphus destructor
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
29-Sep-1999
#accession S66499
#cross-references EMBL:X83875; NID:g999457; PIDN:CAAS8755.1;
#summary Schmidt, M.; Olsson, S.; van der Ploeg, I.; van Hage-Hamster
M.
#journal FEBS Lett. (1995) 370:11-14
#title CDNA analysis of the mite allergen Lep d 1 identifies two
different isoallergens and variants.
#cross-references MUID:95377437
#accession S66499
#molecule_type mRNA
#residues 1-141 #label SCH
#cross-references EMBL:X83875; NID:g999458
CLASSIFICATION #superfamily allergen Der p II

```

```

FEATURE
1-16      #domain signal sequence #status predicted #label SIG
17-141    #product allergen Lep d 1.02 #status predicted #label
MAP
SUMMARY   #length 141 #molecular-weight 14812 #checksum 5305
Query Match      28.6%; Score 69; DB 2; Length 141;
Best Local Similarity 50.0%; Pred. No. 4.71e-01;
Matches 8; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

Db 33 DISGCGS-DTCVIHRG 47
      :::: | | | |
QY 14 ELGGCHGSEPCNIHRG 29

RESULT 13
ENTRY   #type complete
TITLE   homoserine O-acetyltransferase homolog - Haemophilus
ORGANISM influenzae (strain Rd KW20)
#formal_name Haemophilus influenzae
DATE    18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
24-Oct-1997
ACCESSIONS D64113
REFERENCE   A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal   Science (1995) 269:496-512
#title     Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
#cross-references MUID:195350830
#accession D64113
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-358 #label TIGR
#cross-references GB:U32806; GB:L42023; NID:gl574717; PID:gl574721;
TIGR:HL1263
SUMMARY   #length 358 #molecular-weight 39990 #checksum 6643
Query Match      28.6%; Score 69; DB 2; Length 358;
Best Local Similarity 44.4%; Pred. No. 4.71e-01;
Matches 12; Conservative 6; Mismatches 6; Indels 3; Gaps 3;

Db 82 FISSNVLGCKGTGPSSINPQTGPY 108
      ||: |||| | | | |
QY 9 FIGITELGGCHGS-EPCNIH-R-GKPF 32

RESULT 14
ENTRY   #type complete
TITLE   hypothetical protein F7H19.220 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE    23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
23-Jul-1999
ACCESSIONS T05135
REFERENCE   Z15399
#authors   Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.;
Stiekema, W.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.;
Schueller, C.
#submission submitted to the Protein Sequence Database, July 1998
#accession T05135
#molecule_type DNA
#residues   1-502 #label BEV
#cross-references EMBL:AL031018
#experimental_source cultivar Columbia; BAC clone F7H19
GENETICS
#map_position 4
#note         F7H19.220
SUMMARY   #length 502 #molecular-weight 55232 #checksum 5495
Query Match      28.6%; Score 69; DB 2; Length 502;
Best Local Similarity 42.9%; Pred. No. 4.71e-01;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 382 IIGLCELGNCPTTLCGLRG 402
      :::: | | | |
QY 9 FIGITELGGCHGSEPCNIHRG 29

RESULT 15
ENTRY   #type complete
TITLE   hypothetical protein AF0946 - Archaeoglobus fulgidus
ORGANISM #formal_name Archaeoglobus fulgidus
DATE    05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
05-Jun-1998
ACCESSIONS B69368
REFERENCE   A69250
#authors   Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.;
Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman,
J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
C.R.; Venter, J.C.
#journal   Nature (1997) 390:364-370
#title     The complete genome sequence of the hyperthermophilic,
sulfate-reducing archaeon Archaeoglobus fulgidus.
#cross-references MUID:98049343
#accession B69368
#status     preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues   1-1036 #label KLE
#cross-references GB:AE001039; GB:AE000782; NID:g2689362; PID:g2649670;
TIGR:AF0946
SUMMARY   #length 1036 #molecular-weight 112425 #checksum 562
Query Match      28.2%; Score 68; DB 2; Length 1036;
Best Local Similarity 30.0%; Pred. No. 7.20e-01;
Matches 9; Conservative 9; Mismatches 10; Indels 2; Gaps 2;

Db 543 INSEVFK-IDDLAFSGDPDCDV-REKTY 570
      :::: | | | |
QY 3 IKANSFIGITELGGCHGSEPCNIHRGKPF 32

Search completed: Sun Apr 2 18:07:34 2000
Job time : 49 secs.

```

[M][P][E][R][C][H] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:10:13 2000; MasPar time 2.25 Seconds
184.493 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QYIKANSKFITGELGCHGSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 21.282; Variance 66.961; scale 0.318

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Pred. No.	
Result No.	Score	Query Match	Length	ID	Description	
1	122	50.6	25	3	PCT-US95-0	Sequence 19, Applicati
2	122	50.6	25	2	US-08-482-	Sequence 42, Applicati
3	122	50.6	25	2	US-08-478-	Sequence 104, Applicat
4	122	50.6	25	2	US-08-478-	Sequence 42, Applicati
5	122	50.6	61	2	US-08-482-	Sequence 162, Applicat
6	122	50.6	61	2	US-08-478-	Sequence 162, Applicat
7	122	50.6	84	2	US-08-482-	Sequence 161, Applicat
8	122	50.6	84	2	US-08-478-	Sequence 161, Applicat
9	122	50.6	92	2	US-08-482-	Sequence 160, Applicat
10	122	50.6	92	2	US-08-478-	Sequence 160, Applicat
11	122	50.6	129	2	US-08-482-	Sequence 159, Applicat
12	122	50.6	129	1	US-07-945-	Sequence 8, Applicatio
13	122	50.6	129	1	US-08-461-	Sequence 8, Applicatio
14	122	50.6	129	1	US-08-461-	Sequence 8, Applicatio
15	122	50.6	129	1	US-08-462-	Sequence 12, Applicati
16	122	50.6	129	2	US-08-482-	Sequence 8, Applicatio
17	122	50.6	129	1	US-08-462-	Sequence 13, Applicati
18	122	50.6	129	1	US-07-945-	Sequence 12, Applicati
19	122	50.6	129	1	US-08-461-	Sequence 8, Applicatio
20	122	50.6	129	3	PCT-US93-0	Sequence 12, Applicati
21	122	50.6	129	3	PCT-US93-0	Sequence 13, Applicati
22	122	50.6	129	3	PCT-US93-0	Sequence 8, Applicatio
23	122	50.6	129	2	US-08-478-	Sequence 157, Applicat

24	122	50.6	129	1	US-07-945-	Sequence 13, Applicati	6.81e-06
25	122	50.6	129	1	US-08-461-	Sequence 13, Applicati	6.81e-06
26	122	50.6	129	1	US-08-461-	Sequence 12, Applicati	6.81e-06
27	122	50.6	129	2	US-08-482-	Sequence 157, Applicat	6.81e-06
28	122	50.6	129	2	US-08-478-	Sequence 159, Applicat	6.81e-06
29	122	50.6	129	2	US-08-478-	Sequence 8, Applicatio	6.81e-06
30	122	50.6	129	1	US-08-461-	Sequence 13, Applicati	6.81e-06
31	122	50.6	129	1	US-08-461-	Sequence 12, Applicati	6.81e-06
32	122	50.6	130	2	US-08-482-	Sequence 158, Applicat	6.81e-06
33	122	50.6	130	2	US-08-478-	Sequence 158, Applicat	6.81e-06
34	122	50.6	142	1	US-08-288-	Sequence 4, Applicatio	6.81e-06
35	122	50.6	142	2	US-08-910-	Sequence 4, Applicatio	6.81e-06
36	122	50.6	142	2	US-08-910-	Sequence 4, Applicatio	6.81e-06
37	122	50.6	142	2	US-08-905-	Sequence 2, Applicatio	6.81e-06
38	122	50.6	142	2	US-08-905-	Sequence 4, Applicatio	6.81e-06
39	122	50.6	142	1	US-08-288-	Sequence 2, Applicatio	6.81e-06
40	122	50.6	146	1	US-08-462-	Sequence 4, Applicatio	6.81e-06
41	122	50.6	146	1	US-08-461-	Sequence 4, Applicatio	6.81e-06
42	122	50.6	146	3	PCT-US93-0	Sequence 4, Applicatio	6.81e-06
43	122	50.6	146	1	US-07-945-	Sequence 4, Applicatio	6.81e-06
44	122	50.6	146	1	US-08-478-	Sequence 4, Applicatio	6.81e-06
45	122	50.6	146	2	US-08-478-	Sequence 4, Applicatio	6.81e-06

ALIGNMENTS

RESULT 1
ID PCT-US95-04481-19 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX
Sequence 19, Application PC/TUS9504481
CC
XX
Sequence 19, Application PC/TUS9504481
CC
GENERAL INFORMATION:
CC
APPLICANT:
CC
TITLE OF INVENTION: Pharmaceutical Peptide Formulations For Treatment of Du
CC
NUMBER OF SEQUENCES: 54
CC
COMPUTER READABLE FORM:
CC
MEDIUM TYPE: Floppy disk
CC
COMPUTER: IBM PC compatible
CC
OPERATING SYSTEM: PC-DOS/MS-DOS
CC
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CC
CURRENT APPLICATION DATA:
CC
APPLICATION NUMBER: PCT/US95/04481
CC
FILING DATE:
CC
CLASSIFICATION:
CC
PRIOR APPLICATION DATA:
CC
APPLICATION NUMBER: 08/227,772
CC
FILING DATE: April 14, 1994
CC
ATTORNEY/AGENT INFORMATION:
CC
NAME: Vanstone, Darlene A.
CC
REGISTRATION NUMBER: 35,279
CC
REFERENCE/DOCKET NUMBER: 017.5 PCT
CC
TELECOMMUNICATION INFORMATION:
CC
TELEPHONE: (617) 466-6000
CC
TELEFAX: (617) 466-6040
CC
INFORMATION FOR SEQ ID NO: 19:
CC
SEQUENCE CHARACTERISTICS:
CC
LENGTH: 25 amino acids
CC
TYPE: amino acid
CC
STRANDEDNESS:
CC
TOPOLOGY: linear
CC
MOLECULE TYPE: peptide
CC
FRAGMENT TYPE: internal
CC
SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match 50.6%; Score 122; DB 3; Length 25;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DE Sequence 42, Application US/08478572
XX Sequence 42, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 25;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 10 GCHGSEPCIIHRGKPF 25
QY 17 GCHGSEPCNIIHRGKPF 32
RESULT 5
ID US-08-482-142-162 STANDARD; PRT: 61 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 162, Application US/08482142
XX Sequence 162, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard

CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 61 AA; 6669 MW; 17328 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 61;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIIHRGKPF 32
RESULT 6
ID US-08-478-572-162 STANDARD; PRT: 61 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 162, Application US/08478572
XX Sequence 162, Application US/08478572
CC Patent No. 5968526
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean

CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 61 AA; 6669 MW; 17328 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 61;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 7
ID US-08-482-142-161 STANDARD; PRT: 84 AA.
XX
AC xxxxxx
XX
DT
Sequence 161, Application US/08482142
Sequence 161, Application US/08482142
Patent No. 5820862
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:

CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 84;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 8
ID US-08-478-572-161 STANDARD; PRT: 84 AA.
XX
AC xxxxxx
XX
DT
Sequence 161, Application US/08478572
Sequence 161, Application US/08478572
Patent No. 5968526
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC
SQ SEQUENCE 84 AA; 9150 MW; 34277 CN;

Query Match 50.6%; Score 122; DB 2; Length 84;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||

RESULT 9
ID US-08-482-142-160 STANDARD; PRT; 92 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 160, Application US/08482142
XX
XX Sequence 160, Application US/08482142
XX Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142

CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 160:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 92 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC
SQ SEQUENCE 92 AA; 10222 MW; 43773 CN;

Query Match 50.6%; Score 122; DB 2; Length 92;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||

RESULT 10
ID US-08-478-572-160 STANDARD; PRT; 92 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 160, Application US/08478572
XX
XX Sequence 160, Application US/08478572
XX Patent No. 5968536
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC Compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:

CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 160:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 92 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 92 AA; 10222 MW; 43773 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 92;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||||
RESULT 11
ID US-08-482-142-159 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 159, Application US/08482142
XX
XX Sequence 159, Application US/08482142
CC Patent No. 5820862
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/482,142
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.60US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040

CC INFORMATION FOR SEQ ID NO: 159:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 129 AA; 14018 MW; 84629 CN;
SQ
Query Match 50.6%; Score 122; DB 2; Length 129;
Best Local Similarity 93.8%; Pred. No. 6.81e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
||||||| |||||||
RESULT 12
ID US-07-945-288-8 STANDARD; PRT; 129 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 8, Application US/07945288
XX
XX Sequence 8, Application US/07945288
CC Patent No. 5433948
CC GENERAL INFORMATION:
CC APPLICANT: Thomas, Wayne R.
CC APPLICANT: Chua, Kaw-Yan
CC TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
CC TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET, SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/945,288
CC FILING DATE: 19920910
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 580,655
CC FILING DATE: 11 SEPTEMBER 1990
CC APPLICATION NUMBER: 458,642
CC FILING DATE: 13 FEBRUARY 1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 129 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 129 AA; 14022 MW; 84331 CN;
SQ
Query Match 50.6%; Score 122; DB 1; Length 129;
Best Local Similarity 77.8%; Pred. No. 6.81e-06;

CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
 CC TITLE OF INVENTION: DERMATOPHAGOIDES
 CC NUMBER OF SEQUENCES: 13
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: LAHIVE & COCKFIELD
 CC STREET: 60 STATE STREET, SUITE 510
 CC CITY: BOSTON
 CC STATE: MA
 CC COUNTRY: USA
 CC ZIP: 02109
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: ASCII TEXT
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/462,831
 CC FILING DATE:
 CC CLASSIFICATION: 424
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 07/945,288
 CC FILING DATE: 10 SEPTEMBER 1992
 CC APPLICATION NUMBER: US 580,655
 CC FILING DATE: 11 SEPTEMBER 1990
 CC APPLICATION NUMBER: US 458,642
 CC FILING DATE: 13 FEBRUARY 1990
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: MANDRAGOURAS, AMY E.
 CC REGISTRATION NUMBER: 36,207
 CC REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 227-7400
 CC TELEFAX: (617) 227-5941
 CC INFORMATION FOR SEQ ID NO: 12:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 129 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC FEATURE:
 CC NAME/KEY: misc feature
 CC LOCATION: 47
 CC OTHER INFORMATION: /label-xaa is Thr or Ser
 CC FEATURE:
 CC NAME/KEY: misc feature
 CC LOCATION: 114
 CC OTHER INFORMATION: /label-xaa is Asp or Asn
 CC FEATURE:
 CC NAME/KEY: misc feature
 CC LOCATION: 127
 CC OTHER INFORMATION: /label-xaa is Ile or Leu
 CC
 SQ SEQUENCE 129 AA; 14122 MW; 90398 CN;

Query Match 50.6%; Score 122; DB 1; Length 129;
 Best Local Similarity 93.8%; Pred. No. 6.81e-06;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 20 GCHGSEPCIIHRGKPF 35
 | | | | | | | | | |
 QY 17 GCHGSEPCNIHRGKPF 32

Search completed: Sun Apr 2 18:10:17 2000
 Job time : 4 secs.

W P S R L H (TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Apr 2 18:05:02 2000; MasPar time 2.37 Seconds
Tabular output not generated. 218.649 Million cell updates/sec

Title: >US-09-362-731-4
Description: (1-40) from US09362731.pep
Perfect Score: 278
Sequence: 1 PRYKQNTLKLATGKRGPKYKQNTLKLATGKRGVIIGK 40

Scoring table: PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 20.446; Variance 81.810; scale 0.250

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	110	39.6	347	1	US-08-229- Sequence 58, Applicati	5.03e-03
2	110	39.6	347	1	US-08-630- Sequence 58, Applicati	5.03e-03
3	109	39.2	25	2	US-08-484- Sequence 24, Applicati	6.23e-03
4	106	38.1	570	2	US-08-453- Sequence 7, Applicatio	1.18e-02
5	106	38.1	571	2	US-08-453- Sequence 15, Applicati	1.18e-02
6	106	38.1	571	2	US-08-453- Sequence 21, Applicati	1.18e-02
7	90	32.4	13	3	PCT-US94-1 Sequence 20, Applicati	3.39e-01
8	90	32.4	13	3	PCT-US93-0 Sequence 24, Applicati	3.39e-01
9	90	32.4	13	2	US-08-480- Sequence 24, Applicati	3.39e-01
10	90	32.4	13	2	US-08-596- Sequence 164, Applicat	3.39e-01
11	90	32.4	13	1	US-08-305- Sequence 1, Applicatio	3.39e-01
12	90	32.4	13	1	US-08-465- Sequence 22, Applicati	3.39e-01
13	90	32.4	13	3	PCT-US92-0 Sequence 18, Applicati	3.39e-01
14	90	32.4	13	1	US-07-831- Sequence 5, Applicatio	3.39e-01
15	90	32.4	13	2	US-08-596- Sequence 8, Applicatio	3.39e-01
16	90	32.4	13	2	US-08-488- Sequence 24, Applicati	3.39e-01
17	90	32.4	13	3	PCT-US95-0 Sequence 53, Applicati	3.39e-01
18	90	32.4	13	3	PCT-US95-0 Sequence 99, Applicati	3.39e-01
19	90	32.4	13	3	PCT-US95-0 Sequence 8, Applicatio	3.39e-01
20	86	30.9	13	3	PCT-US95-0 Sequence 59, Applicati	7.71e-01
21	85	30.6	13	1	US-08-619- Sequence 9, Applicatio	9.45e-01
22	85	30.6	13	2	US-08-634- Sequence 9, Applicatio	9.45e-01
23	83	29.9	13	3	PCT-US95-0 Sequence 58, Applicati	1.42e-00

ALIGNMENTS

RESULT 1
ID US-08-229-781-58 STANDARD; PRT; 347 AA.

AC xxxxxx

XX

DT

XX

DE

XX

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

79 28.4 13 3 PCT-US95-0 Sequence 57, Applicati
75 27.0 348 1 US-08-229- Sequence 50, Applicati
75 27.0 348 1 US-08-630- Sequence 50, Applicati
70 25.2 1148 2 US-08-313- Sequence 58, Applicati
68 24.5 572 2 US-08-453- Sequence 9, Applicatio
67 24.1 1082 1 US-08-429- Sequence 2, Applicatio
67 24.1 1082 1 US-08-106- Sequence 2, Applicatio
67 24.1 1139 1 US-08-832- Sequence 2, Applicatio
67 24.1 1139 1 US-08-832- Sequence 2, Applicatio
66 23.7 364 3 PCT-US96-1 Sequence 2, Applicatio
66 23.7 364 3 PCT-US96-1 Sequence 2, Applicatio
65 23.4 393 3 PCT-US96-1 Sequence 3, Applicatio
65 23.4 446 2 US-08-244- Sequence 5, Applicatio
65 23.4 446 2 US-08-244- Sequence 5, Applicatio
64 23.0 977 2 US-08-673- Sequence 10, Applicati
63 22.7 976 2 US-08-702- Sequence 8, Applicatio
63 22.7 976 2 US-08-702- Sequence 18, Applicati
63 22.7 976 2 PCT-US95-0 Sequence 18, Applicati
63 22.7 976 2 US-08-449- Sequence 18, Applicati
63 22.7 1147 2 US-08-668- Sequence 38, Applicati
63 22.7 1147 1 US-08-131- Sequence 38, Applicati
62 22.3 120 1 US-08-336- Sequence 17, Applicati
62 22.3 554 3 PCT-US94-0 Sequence 7, Applicatio
45 22.3 554 3 PCT-US94-0 Sequence 7, Applicatio

Sequence 58, Application US/08229781

Sequence 58, Application US/08229781

Patent No. 5589174

GENERAL INFORMATION:

APPLICANT: Yoshinobu OKUNO et al.

TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY

NUMBER OF SEQUENCES: 58

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/229,781

FILING DATE: April 19, 1994

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/054,016

FILING DATE: April 29, 1993

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELEPHONE: 202-371-8850

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 347 amino acids

TYPE: amino acid

STRANDEDNESS: single

CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 1; Length 347;
Best Local Similarity 63.3%; Pred. No. 5.03e-03;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKONTLKLATGMNV 120
|||::| |||||
QY 6 QNTLKLATGKRGPKYVKONTLKLATGKGV 35

RESULT 2
ID US-08-630-918-58 STANDARD; PRT; 347 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 58, Application US/08630918
XX
CC Sequence 58, Application US/08630918
CC Patent No. 5631350
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/630,918
CC FILING DATE: April 5, 1996
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 1; Length 347;
Best Local Similarity 63.3%; Pred. No. 5.03e-03;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKONTLKLATGMNV 120
|||::| |||||
QY 6 QNTLKLATGKRGPKYVKONTLKLATGKGV 35

RESULT 3
ID US-08-484-905-24 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 24, Application US/08484905
XX
CC Sequence 24, Application US/08484905
CC Patent No. 5976551
CC GENERAL INFORMATION:
CC APPLICANT: Mottez, Estelle
CC APPLICANT: Abastado, Jean-Pierre
CC APPLICANT: Kourilsky, Philippe
CC TITLE OF INVENTION: An Altered Major Histocompatibility
CC TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the
CC TITLE OF INVENTION: Determinant
CC NUMBER OF SEQUENCES: 127
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy Disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,905

```

CC FILING DATE: 13-SEPT-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MGS101CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-873-8794
CC TELEFAX: (404)-873-8795
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 570 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Influenza virus
CC INDIVIDUAL ISOLATE: A/Beijing/32/92 RHA
CC FEATURE:
CC NAME/KEY: ACPV 61K protein signal sequence
CC LOCATION: 1 to 18
CC FEATURE:
CC NAME/KEY: mature rHA
CC LOCATION: 19 to 552
CC
CC SQ SEQUENCE 570 AA; 64054 MW; 1596768 CN;
Query Match 38.1%; Score 106; DB 2; Length 570;
Best Local Similarity 77.8%; Pred. No. 1.18e-02;
Matches 14; Conservative 3; Mismatches 1; Indels

Db 326 PRYVKQNTLKLATGMNV 343
Qy 18 PRYVKQNTLKLATGKKV 35
I:|||||||:|
|:|||||||:|

RESULT 5 STANDARD; PRT; 571 AA.
XX US-08-453-848-15
XX AC xxxxxx
XX
XX DT
XX DE
XX Sequence 15, Application US/08453848
XX
XX Sequence 15, Application US/08453848
XX Patent No. 5858368
XX GENERAL INFORMATION:
XX APPLICANT: Smith, Gale Eugene
XX APPLICANT: Volnovitz, Franklin
XX APPLICANT: Wilkonson, Bethanie Eident
XX APPLICANT: Voznesensky, Andrei I.
XX APPLICANT: Hackett, Craig Stanley
XX TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
XX TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
XX NUMBER OF SEQUENCES: 31
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Patrea L. Pabst
XX STREET: 2800 One Atlantic Center
XX STREET: 1201 West Peachtree Street
XX CITY: Atlanta
XX STATE: GA
XX COUNTRY: USA
XX ZIP: 30309-3450
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: PatentIn Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:

```

CC APPLICATION NUMBER: US/08/453,848
CC FILING DATE: 30-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/120,607
CC FILING DATE: 13-SEPT-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MGS101CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-873-8794
CC TELEFAX: (404)-873-8795
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 571 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Influenza virus
CC INDIVIDUAL ISOLATE: A/Shandong/9/93 rHA
CC FEATURE:
CC NAME/KEY: AcNPV 61K protein signal sequence
CC LOCATION: 1 to 18
CC NAME/KEY: mature rHA
CC LOCATION: 19 to 553
CC SEQUENCE 571 AA; 64061 MW; 1600797 CN;
SQ
Query Match 38.1%; Score 106; DB 2; Length 571;
Best Local Similarity 77.8%; Pred. No. 1.18e-02;
Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 327 PRYVKQNTLKLATGMNV 344
QY 18 PRYVKQNTLKLATGKGV 35
RESULT 6
ID US-08-453-848-21 STANDARD; PRT; 571 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 21, Application US/08453848
XX Sequence 21, Application US/08453848
CC Patent No. 5858368
CC GENERAL INFORMATION:
CC APPLICANT: Smith, Gale Eugene
CC APPLICANT: Volnovitz, Franklin
CC APPLICANT: Wilkinson, Bethanie Eident
CC APPLICANT: Voznesensky, Andrei I.
CC APPLICANT: Hackett, Craig Stanway
CC TITLE OF INVENTION: A METHOD FOR PRODUCING INFLUENZA
CC TITLE OF INVENTION: HEMAGGLUTININ MULTIVALENT VACCINES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Patrea L. Pabst
CC STREET: 2800 One Atlantic Center
CC STREET: 1201 West Peachtree Street
CC CITY: Atlanta
CC STATE: GA
CC COUNTRY: USA
CC ZIP: 30309-3450
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/453,848
CC FILING DATE: 30-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/120,607
CC FILING DATE: 13-SEPT-1993
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: MGS101CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404)-873-8794
CC TELEFAX: (404)-873-8795
CC INFORMATION FOR SEQ ID NO: 21:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 571 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Influenza virus
CC INDIVIDUAL ISOLATE: A/Johannesburg/33/94 rHA
CC FEATURE:
CC NAME/KEY: AcNPV 61K protein signal sequence
CC LOCATION: 1 to 18
CC NAME/KEY: mature rHA
CC LOCATION: 19 to 569
CC SEQUENCE 571 AA; 64271 MW; 1597291 CN;
SQ
Query Match 38.1%; Score 106; DB 2; Length 571;
Best Local Similarity 77.8%; Pred. No. 1.18e-02;
Matches 14; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Db 327 PRYVKQNTLKLATGMNV 344
QY 18 PRYVKQNTLKLATGKGV 35
RESULT 7
ID PCT-US94-10257A-20 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 20, Application PC/TUS9410257A
XX Sequence 20, Application PC/TUS9410257A
CC GENERAL INFORMATION:
CC APPLICANT: IMMULOGIC PHARMACEUTICAL CORPORATION
CC APPLICANT: BRIGITTE DEVAUX
CC APPLICANT: JONATHAN B. ROTHBARD
CC APPLICANT: DAWN SMILER
CC TITLE OF INVENTION: USES OF MYELIN OLIGODENDROCYTE
CC TITLE OF INVENTION: GLYCOPROTEIN AND PEPTIDE PORTIONS THEREOF IN PROTOCOLS
CC TITLE OF INVENTION: RELATED TO AUTOIMMUNE DISEASE
CC NUMBER OF SEQUENCES: 95
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: Massachusetts

CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC

Query Match 100.0%: score

Query Match 100.0%; Score 106; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 11
ID US-08-478-572-50 STANDARD; PRT; 26 AA.

XX
AC
XX
DT

Sequence 50, Application US/08478572

Sequence 50, Application US/08478572

Patent No. 5968526

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/478,572

FILING DATE: 07-June-1995

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/445,307

FILING DATE:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017.6US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 50:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

SEQUENCE 26 AA; 2818 MW; 3780 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 12
ID US-08-482-142-170 STANDARD; PRT; 26 AA.

XX
AC
XX
DT

Sequence 170, Application US/08482142

Sequence 170, Application US/08482142

Patent No. 5820862

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/482,142

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/445,307

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

REFERENCE/DOCKET NUMBER: 017.6US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 466-6000

TELEFAX: (617) 466-6040

INFORMATION FOR SEQ ID NO: 170:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

SEQUENCE 26 AA; 2778 MW; 3598 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
XXXXXXXXXXXX
Qy 1 HEIKKVLVPGCHGS 14

RESULT 13
ID US-08-482-142-50 STANDARD; PRT; 26 AA.
XX
AC


```

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/478,572
CC FILING DATE: 07-June-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 170:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 26 AA: 2778 MW; 3598 CN;

Query Match 100.0%; Score 106; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. NO. 2.32e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14

```

Search completed: Sun Apr 2 17:54:32 2000
Job time : 4 secs.

CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: AMINO ACIDS
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 52.2%; Score 117; DB 3; Length 26;
Best Local Similarity 63.8%; Pred. No. 2.84e-04;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 3 VDVKDSANHEIKKVLPGCHGS 24
: : : : :
QY 10 IGITELGGHEIKKVLPGCHGS 31

RESULT 15
ID US-08-484-296-52 STANDARD; PRT; 26 AA.
AC xxxxxx
XX
XX
XX
XX
XX
XX
XX

Sequence 52, Application US/08484296

Sequence 52, Application US/08484296

GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 52:
CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 26 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 26 AA; 2802 MW; 3868 CN;

Query Match 52.2%; Score 117; DB 9; Length 26;
Best Local Similarity 63.6%; Pred. No. 2.84e-04;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 3 VDVKDSANHEIKKVLPGCHGS 24
: : : : :
QY 10 IGITELGGHEIKKVLPGCHGS 31

Search completed: Sun Apr 2 17:50:05 2000
Job time : 14 secs.

CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Cancer
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488,351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446,692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229,275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057,166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 52.7%; Score 118; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITELGG 37
Qy 1 QYKANSKFIGITELGG 17

RESULT 13
ID US-07-881-396A-51 STANDARD; PRT: 26 AA.
XX xxxxxx
AC xxxxxx
XX xxxxxx
DT xxxxxx
DE xxxxxx
XX Sequence 51, Application US/07881396A
CC Sequence 51, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.

CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 26 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 26 AA; 2844 MW; 3796 CN;

Query Match 52.2%; Score 117; DB 3; Length 26;
Best Local Similarity 63.6%; Pred. No. 2.84e-04;
Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 3 VDVKDEANHEIKKVLVPGCHGS 24
Qy 10 IGITELGGHEIKKVLVPGCHGS 31

RESULT 14
ID US-07-881-396A-52 STANDARD; PRT: 26 AA.
XX xxxxxx
AC xxxxxx
XX xxxxxx
DT xxxxxx
DE xxxxxx
XX Sequence 52, Application US/07881396A
CC Sequence 52, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 52.7%; Score 118; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITELGG 37
QY 1 QYKANSKFIGITELGG 17

RESULT 10
ID PCT-US94-04832A-35 STANDARD; PRT; 47 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 35, Application PC/TUS9404832A
GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic LHRH peptide constructs
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: 400 Garden City Plaza
STREET: Garden City
CITY: Garden City
STATE: NY
COUNTRY: US
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04832A
FILING DATE: 13-APR-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 47 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 52.7%; Score 118; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITELGG 37
QY 1 QYKANSKFIGITELGG 17

RESULT 10
ID PCT-US94-04832A-35 STANDARD; PRT; 47 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

RESULT 11
ID US-08-229-275-35 STANDARD; PRT; 47 AA.
XX
AC xxxxxx
XX
XX
DT
XX

Sequence 35, Application US/08229275

Sequence 35, Application US/08229275

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs as

TITLE OF INVENTION: vaccines for treatment of prostate cancer and induction

TITLE OF INVENTION: infertility

NUMBER OF SEQUENCES: 53

CORRESPONDENCE ADDRESS:

ADDRESSEE: M. Lisa Wilson

STREET: 25 Davids Drive

CITY: Hauppauge

STATE: NY

COUNTRY: US

ZIP: 11788

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/229,275

FILING DATE: 13-APR-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Wilson, M L

REGISTRATION NUMBER: 34,045

REFERENCE/DOCKET NUMBER: 2003Z

TELEPHONE: (516)273-2828

TELEFAX: (516)273-1717

INFORMATION FOR SEQ ID NO: 35:

SEQUENCE CHARACTERISTICS:

LENGTH: 47 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 52.7%; Score 118; DB 6; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYKANSKFIGITELGG 37

QY 1 QYKANSKFIGITELGG 17

RESULT 12

ID US-08-488-320A-35 STANDARD; PRT; 47 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 35, Application US/08488320A

Sequence 35, Application US/08488320A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH

CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 52.7%; Score 118; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITELGG 21
QY 1 QYIKANSKFIGITELGG 17

RESULT 8
ID US-08-488-320A-63 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT

Sequence 63, Application US/08488320A

Sequence 63, Application US/08488320A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH

TITLE OF INVENTION: And A Helper T-cell Epitope For Treatment Of Prostate Cancer

TITLE OF INVENTION: And Induction Of Infertility

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,320A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/488,351

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/446,692

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/229,275

FILING DATE: 14-APR-1994

APPLICATION NUMBER: 08/057,166

FILING DATE: 27-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-41460S4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)758-4800

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO: 57:

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 63:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 37 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 52.7%; Score 118; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITELGG 21
QY 1 QYIKANSKFIGITELGG 17

RESULT 9
ID US-08-488-320A-57 STANDARD; PRT; 37 AA.
XX
AC xxxxxx
XX
DT

Sequence 57, Application US/08488320A

Sequence 57, Application US/08488320A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH

TITLE OF INVENTION: And A Helper T-cell Epitope For Treatment Of Prostate Cancer

TITLE OF INVENTION: And Induction Of Infertility

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: U.S.A.

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,320A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/488,351

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/446,692

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: 08/229,275

FILING DATE: 14-APR-1994

APPLICATION NUMBER: 08/057,166

FILING DATE: 27-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Maria C. H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-41460S4

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)758-4800

TELEFAX: (212)751-6849

INFORMATION FOR SEQ ID NO: 57:

CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 29 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 29 AA; 3195 MW; 5252 CN;

Query Match 52.7%; Score 118; DB 12; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITELGG 19
|||||
QY 1 QYIKANSKFIGITELGG 17

RESULT 5
ID US-09-362-731-5 STANDARD; PRT; 32 AA.

AC xxxxxx

XX

DT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 5, Application US/09362731

Sequence 5, Application US/09362731

GENERAL INFORMATION:

APPLICANT: U.C.B. S.A.

TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

FILE REFERENCE: P.OCB.09/WO

CURRENT APPLICATION NUMBER: US/09/362.731

CURRENT FILING DATE: 1999-07-29

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 32

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: peptide

SEQUENCE 32 AA; 3503 MW; 5050 CN;

Query Match 52.7%; Score 118; DB 18; Length 32;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITELGG 17
|||||
QY 1 QYIKANSKFIGITELGG 17

RESULT 6
ID PCT-US94-04832A-57 STANDARD; PRT; 37 AA.

AC xxxxxx

XX

DT

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 57, Application PC/TUS9404832A

Sequence 57, Application PC/TUS9404832A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 52.7%; Score 118; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITELGG 37
|||||
QY 1 QYIKANSKFIGITELGG 17

RESULT 7
ID PCT-US94-04832A-63 STANDARD; PRT; 37 AA.

AC xxxxxx

XX

DT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 63, Application PC/TUS9404832A

Sequence 63, Application PC/TUS9404832A

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE:

STREET: 400 Garden City Plaza

CITY: Garden City

STATE: NY

COUNTRY: US

ZIP: 11530

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/04832A

DE Sequence 3, Application US/09362731
XX
CC Sequence 3, Application US/09362731
CC GENERAL INFORMATION:
CC APPLICANT: U.C.B. S.A.
CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE
CC FILE OF INVENTION: TREATMENT OF ALLERGY
CC FILE REFERENCE: P.UCB.09/WO
CC CURRENT APPLICATION NUMBER: US/09/362,731
CC CURRENT FILING DATE: 1999-07-29
CC NUMBER OF SEQ ID NOS: 17
CC SOFTWARE: Patent in Ver. 2.1
CC SEQ ID NO 3
CC LENGTH: 137
CC TYPE: PRT
CC ORGANISM: Artificial Sequence
CC FEATURE:
CC OTHER INFORMATION: Description of Artificial Sequence: synthetic
CC OTHER INFORMATION: peptide
CC SEQUENCE 137 AA; 14747 MW; 93695 CN;
SQ

Query Match 58.0%; Score 130; DB 18; Length 137;
Best Local Similarity 90.5%; Pred. No. 7.11e-06;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Db 2 QYIKANSKFIGITELGGQYIK 22
QY 1 QYIKANSKFIGITELGGHEIK 21

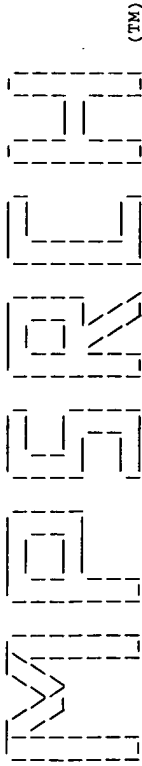
RESULT 3
ID US-08-328-912B-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08328912B
XX
CC Sequence 37, Application US/08328912B
CC GENERAL INFORMATION:
CC APPLICANT: Wang, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 61
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/328,912B
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC

CC Sequence 3, Application US/08718490A
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08718490A
XX
CC Sequence 37, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 29 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 29 AA; 3195 MW; 5252 CN;
SQ

Query Match 52.7%; Score 118; DB 7; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.15e-04;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFIGITELGG 19
QY 1 QYIKANSKFIGITELGG 17

RESULT 4
ID US-08-718-490A-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 37, Application US/08718490A
XX
CC Sequence 37, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang Y1
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:49:51 2000; MasPar time 12.10 Seconds
273.319 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-1
Description: (1-31) from US09362731.pap
Perfect Score: 224
Sequence: 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

Scoring table: PAM 150
Gap 11

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1:P9 2:U060 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 25.291; Variance 73.379; scale 0.345

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	224	100.0	31	18	US-09-362-Sequence 1, Applicatio	3.72e-18
2	130	58.0	137	18	US-09-362-Sequence 3, Applicatio	7.11e-06
3	118	52.7	29	7	US-08-328-Sequence 37, Applicati	2.15e-04
4	118	52.7	29	12	US-08-718-Sequence 37, Applicati	2.15e-04
5	118	52.7	32	18	US-09-362-Sequence 5, Applicatio	2.15e-04
6	118	52.7	37	1	PCT-US94-0 Sequence 57, Applicati	2.15e-04
7	118	52.7	37	1	PCT-US94-0 Sequence 63, Applicati	2.15e-04
8	118	52.7	37	9	US-08-488-Sequence 57, Applicati	2.15e-04
9	118	52.7	37	9	US-08-488-Sequence 57, Applicati	2.15e-04
10	118	52.7	47	1	PCT-US94-0 Sequence 35, Applicati	2.15e-04
11	118	52.7	47	6	US-08-229-Sequence 35, Applicati	2.15e-04
12	118	52.7	47	9	US-08-488-Sequence 35, Applicati	2.15e-04
13	117	52.2	26	3	US-07-881-Sequence 51, Applicati	2.84e-04
14	117	52.2	26	3	US-07-881-Sequence 52, Applicati	2.84e-04
15	117	52.2	26	9	US-08-484-Sequence 52, Applicati	2.84e-04
16	117	52.2	26	6	US-08-227-Sequence 51, Applicati	2.84e-04
17	117	52.2	26	6	US-08-445-Sequence 52, Applicati	2.84e-04
18	117	52.2	26	6	US-08-227-Sequence 52, Applicati	2.84e-04
19	117	52.2	26	8	US-08-445-Sequence 51, Applicati	2.84e-04
20	117	52.2	26	9	US-08-484-Sequence 51, Applicati	2.84e-04

21	116	51.8	25	15	US-09-060-Sequence 37, Applicati	3.77e-04
22	116	51.8	25	15	US-09-060-Sequence 37, Applicati	3.77e-04
23	116	51.8	158	15	US-09-060-Sequence 8, Applicatio	3.77e-04
24	116	51.8	158	15	US-09-060-Sequence 8, Applicatio	3.77e-04
25	113	50.4	25	6	US-08-227-Sequence 169, Applicat	8.70e-04
26	113	50.4	25	15	US-09-060-Sequence 36, Applicati	8.70e-04
27	113	50.4	25	15	US-09-060-Sequence 36, Applicati	8.70e-04
28	113	50.4	26	6	US-08-227-Sequence 50, Applicati	8.70e-04
29	113	50.4	26	9	US-08-484-Sequence 50, Applicati	8.70e-04
30	113	50.4	26	8	US-08-445-Sequence 50, Applicati	8.70e-04
31	113	50.4	26	6	US-08-445-Sequence 170, Applicat	8.70e-04
32	113	50.4	26	6	US-08-227-Sequence 170, Applicat	8.70e-04
33	113	50.4	26	3	US-07-881-Sequence 50, Applicati	8.70e-04
34	113	50.4	26	9	US-08-484-Sequence 170, Applicat	8.70e-04
35	113	50.4	27	6	US-08-227-Sequence 171, Applicat	8.70e-04
36	113	50.4	27	8	US-08-445-Sequence 171, Applicat	8.70e-04
37	113	50.4	27	9	US-08-484-Sequence 171, Applicat	8.70e-04
38	113	50.4	35	9	US-08-460-Sequence 7, Applicatio	8.70e-04
39	113	50.4	145	9	US-08-460-Sequence 6, Applicatio	8.70e-04
40	113	50.4	146	6	US-08-227-Sequence 4, Applicatio	8.70e-04
41	113	50.4	146	9	US-08-484-Sequence 4, Applicatio	8.70e-04
42	113	50.4	146	8	US-08-445-Sequence 4, Applicatio	8.70e-04
43	113	50.4	146	3	US-07-881-Sequence 4, Applicatio	8.70e-04
44	113	50.4	158	15	US-09-060-Sequence 6, Applicatio	8.70e-04
45	113	50.4	158	15	US-09-060-Sequence 6, Applicatio	8.70e-04

ALIGNMENTS

RESULT 1 US-09-362-731-1 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
XX
DT

Sequence 1, Application US/09362731

Sequence 1, Application US/09362731

GENERAL INFORMATION:

APPLICANT: U.C.B. S.A.

TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

TITLE OF INVENTION: TREATMENT OF ALLERGY

FILE REFERENCE: P.UCB.09/WO

CURRENT APPLICATION NUMBER: US/09/362,731

CURRENT FILING DATE: 1999-07-29

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1

LENGTH: 31

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: peptide

SEQUENCE 31 AA; 3325 MW; 5513 CN;

Query Match 100.0%; Score 224; DB 18; Length 31;

Best Local Similarity 100.0%; Pred. No. 3.72e-18;

Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

|||||

QY 1 QYIKANSKFIGITELGGHEIKKVLPGCHGS 31

RESULT 2

ID US-09-362-731-3 STANDARD; PRT; 137 AA.

XX

AC xxxxxx

XX

DT

XX

M P E R L F
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 17:54:49 2000; MasPar time 10.17 Seconds
146.872 Million cell updates/sec

Tabular output not generated.

Title: >US-09-362-731-2
Description: (1-14) from US09362731.pep
Perfect Score: 106
Sequence: 1 HEIKKVLVPGCHGS 14

Scoring table: PAM 150
Gap 15

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:060 3:07 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU6 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 20.674; Variance 49.657; scale 0.416

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	106	100.0	14	18	US-09-362-Sequence 2, Applicatio	1.03e-05
2	106	100.0	16	9	US-08-484-Sequence 59, Applicati	1.03e-05
3	106	100.0	16	6	US-08-227-Sequence 59, Applicati	1.03e-05
4	106	100.0	16	8	US-08-445-Sequence 59, Applicati	1.03e-05
5	106	100.0	16	3	US-07-881-Sequence 59, Applicati	1.03e-05
6	106	100.0	25	3	US-07-881-Sequence 42, Applicati	1.03e-05
7	106	100.0	25	6	US-08-227-Sequence 169, Applicati	1.03e-05
8	106	100.0	25	8	US-08-445-Sequence 42, Applicati	1.03e-05
9	106	100.0	25	8	US-08-445-Sequence 169, Applicati	1.03e-05
10	106	100.0	25	9	US-08-484-Sequence 169, Applicati	1.03e-05
11	106	100.0	25	6	US-08-227-Sequence 42, Applicati	1.03e-05
12	106	100.0	25	9	US-08-484-Sequence 42, Applicati	1.03e-05
13	106	100.0	26	6	US-08-227-Sequence 51, Applicati	1.03e-05
14	106	100.0	26	3	US-07-881-Sequence 50, Applicati	1.03e-05
15	106	100.0	26	3	US-07-881-Sequence 51, Applicati	1.03e-05
16	106	100.0	26	8	US-08-445-Sequence 170, Applicati	1.03e-05
17	106	100.0	26	9	US-08-484-Sequence 51, Applicati	1.03e-05
18	106	100.0	26	6	US-08-227-Sequence 170, Applicati	1.03e-05
19	106	100.0	26	8	US-08-445-Sequence 50, Applicati	1.03e-05
20	106	100.0	26	9	US-08-484-Sequence 52, Applicati	1.03e-05

21	106	100.0	26	6	US-08-227-Sequence 52, Applicati	1.03e-05
22	106	100.0	26	3	US-07-881-Sequence 50, Applicati	1.03e-05
23	106	100.0	26	9	US-08-484-Sequence 50, Applicati	1.03e-05
24	106	100.0	26	8	US-08-445-Sequence 51, Applicati	1.03e-05
25	106	100.0	26	6	US-08-227-Sequence 50, Applicati	1.03e-05
26	106	100.0	26	9	US-08-484-Sequence 170, Applicati	1.03e-05
27	106	100.0	26	8	US-08-445-Sequence 52, Applicati	1.03e-05
28	106	100.0	27	6	US-08-227-Sequence 171, Applicati	1.03e-05
29	106	100.0	27	8	US-08-484-Sequence 171, Applicati	1.03e-05
30	106	100.0	27	9	US-08-445-Sequence 171, Applicati	1.03e-05
31	106	100.0	31	18	US-09-362-Sequence 1, Applicatio	1.03e-05
32	106	100.0	35	9	US-08-460-Sequence 7, Applicatio	1.03e-05
33	106	100.0	145	9	US-08-460-Sequence 6, Applicatio	1.03e-05
34	106	100.0	146	8	US-08-445-Sequence 4, Applicatio	1.03e-05
35	106	100.0	146	6	US-08-227-Sequence 4, Applicatio	1.03e-05
36	106	100.0	146	3	US-07-881-Sequence 4, Applicatio	1.03e-05
37	106	100.0	146	9	US-08-484-Sequence 4, Applicatio	1.03e-05
38	95	89.6	61	8	US-08-445-Sequence 162, Applicat	4.47e-04
39	95	89.6	61	6	US-08-227-Sequence 162, Applicat	4.47e-04
40	95	89.6	84	6	US-08-227-Sequence 161, Applicat	4.47e-04
41	95	89.6	129	6	US-08-227-Sequence 159, Applicat	4.47e-04
42	95	89.6	129	6	US-08-227-Sequence 157, Applicat	4.47e-04
43	95	89.6	129	8	US-08-445-Sequence 159, Applicat	4.47e-04
44	95	89.6	130	8	US-08-445-Sequence 158, Applicat	4.47e-04
45	95	89.6	130	6	US-08-227-Sequence 158, Applicat	4.47e-04

ALIGNMENTS

RESULT 1
ID US-09-362-731-2 STANDARD; PRT; 14 AA.

XX xxxxxx

DT

XX Sequence 2, Application US/09362731

DE Sequence 2, Application US/09362731

CC Sequence 2, Application US/09362731

CC GENERAL INFORMATION:

CC APPLICANT: U.C.B. S.A.

CC TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

CC FILE REFERENCE: P.UGB.09/WO

CC CURRENT APPLICATION NUMBER: US/09/362,731

CC CURRENT FILING DATE: 1999-07-29

CC NUMBER OF SEQ ID NOS: 17

CC SOFTWARE: PatentIn Ver. 2.1

CC SEQ ID NO 2

CC LENGTH: 14

CC TYPE: PRT

CC ORGANISM: Artificial Sequence

CC FEATURE:

CC OTHER INFORMATION: Description of Artificial Sequence: synthetic

CC OTHER INFORMATION: peptide

CC SEQUENCE 14 AA; 1504 MW; 1224 CN;

Query Match 100.0%; Score 106; DB 18; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
1 HEIKKVLVPGCHGS 14

QY 1 HEIKKVLVPGCHGS 14

RESULT 2
ID US-08-484-296-59 STANDARD; PRT; 16 AA.

XX xxxxxx

DT

```

DE Sequence 59, Application US/08484296
XX
CC Sequence 59, Application US/08484296
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 59:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 16 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 16 AA; 1730 MW; 1569 CN;

Query Match 100.0%; Score 106; DB 9; Length 16;
Best Local Similarity 100.0%; Pred.No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
| | | | | | | | | |
QY 1 HEIKKVLVPGCHGS 14

RESULT 3
ID US-08-227-772-59 STANDARD; PRT; 16 AA.
XX
XX xxxxxx
XX
XX
XX
XX
XX
DE Sequence 59, Application US/08227772
XX
CC Sequence 59, Application US/08227772
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia

```

CC STATE: MA
 CC COUNTRY: USA
 CC ZIP: 02154
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: ASCII TEXT
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/445,307A
 CC FILING DATE: 07 June 1995
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: CRAIG, ANNE I.
 CC REGISTRATION NUMBER: 32,976
 CC REFERENCE/DOCKET NUMBER: 017.605
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 466-6000
 CC TELEFAX: (617) 466-6040
 CC INFORMATION FOR SEQ ID NO: 59:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 16 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC FRAGMENT TYPE: N-terminal
 CC SEQUENCE 16 AA; 1730 MW; 1569 CN;
 SQ

Query Match 100.0%; Score 106; DB 8; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPCGCHS 14
 QY 1 HEIKKVLVPCGCHS 14

RESULT 5
 ID US-07-881-396A-59 STANDARD; PRT; 16 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX

Sequence 59, Application US/07881396A
 GENERAL INFORMATION:
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 119
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 STATE STREET SUITE 510
 CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207
 REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 25 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207

CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (617) 227-7400
 CC TELEFAX: (617) 227-5941
 CC INFORMATION FOR SEQ ID NO: 59:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 16 amino acids
 CC TYPE: AMINO ACID
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC FRAGMENT TYPE: N-terminal
 CC SEQUENCE 16 AA; 1730 MW; 1569 CN;
 SQ

Query Match 100.0%; Score 106; DB 3; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPCGCHS 14
 QY 1 HEIKKVLVPCGCHS 14

RESULT 6
 ID US-07-881-396A-42 STANDARD; PRT; 25 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX

Sequence 42, Application US/07881396A
 GENERAL INFORMATION:
 APPLICANT: Kuo, Mei-chang
 APPLICANT: Garman, Richard
 APPLICANT: Greenstein, Julia
 TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
 TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
 NUMBER OF SEQUENCES: 119
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD
 STREET: 60 STATE STREET SUITE 510
 CITY: BOSTON
 STATE: MA
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII TEXT
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207
 REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-7400
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 42:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 25 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 APPLICATION NUMBER: US/07/881,396A
 FILING DATE: 19920508
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: MANDRAGOURAS, AMY E.
 REGISTRATION NUMBER: P36,207

Query Match 100.0%; Score 106; DB 3; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.03e-05;

CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 169:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 25 AA; 2721 MW; 3390 CN;
SQ
Query Match 100.0%; Score 106; DB 8; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
RESULT 10
ID US-08-484-296-169 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 169, Application US/08484296
CC
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET

CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.6US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 169:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 25 AA; 2721 MW; 3390 CN;
SQ
Query Match 100.0%; Score 106; DB 9; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
RESULT 11
ID US-08-227-772-42 STANDARD; PRT: 25 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 42, Application US/08227772
CC
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 197
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match 100.0%; Score 106; DB 6; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14

RESULT 12
ID US-08-484-296-42 STANDARD; PRT: 25 AA.

XX xxxxxx

AC

XX

XX

DT

XX

DE

XX

Sequence 42, Application US/08484296

Sequence 42, Application US/08484296

GENERAL INFORMATION:

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

APPLICANT: Kuo, Mei-chang

APPLICANT: Rogers, Bruce

APPLICANT: Franzen, Henry

APPLICANT: Chen, Xian

APPLICANT: Evans, Sean

APPLICANT: Shaked, Ze'ev

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 207

CORRESPONDENCE ADDRESS:

ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION

STREET: 610 LINCOLN STREET

CITY: WALTHAM

STATE: MA

COUNTRY: USA

ZIP: 02154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,296

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/445,307

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: CRAIG, ANNE I.

REGISTRATION NUMBER: 32,976

CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match 100.0%; Score 106; DB 9; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 HEIKKVLVPGCHGS 14
QY 1 HEIKKVLVPGCHGS 14

RESULT 13
ID US-08-227-772-51 STANDARD; PRT: 26 AA.

XX xxxxxx

AC

XX

XX

DT

XX

DE

XX

Sequence 51, Application US/08227772

Sequence 51, Application US/08227772

GENERAL INFORMATION:

APPLICANT: Kuo, Mei-chang

APPLICANT: Garman, Richard

APPLICANT: Greenstein, Julia

TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS

TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)

NUMBER OF SEQUENCES: 197

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 60 STATE STREET SUITE 510

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/227,772

FILING DATE: 14-APR-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/881,396

FILING DATE: 08-MAY-1992

ATTORNEY/AGENT INFORMATION:

NAME: MANDRAGOURAS, AMY E.

REGISTRATION NUMBER: P36,207

REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:

LENGTH: 26 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: N-terminal

SEQUENCE 26 AA; 2844 MW; 3796 CN;

Query Match 100.0%; Score 106; DB 6; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
|||||

RESULT 14
ID US-07-881-396A-50 STANDARD; PRT; 26 AA.

XX AC xxxxxx

Sequence 50, Application US/07881396A

Sequence 50, Application US/07881396A

GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/881,396A
FILING DATE: 19920508
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-7400

INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
SEQUENCE 26 AA; 2818 MW; 3780 CN;

Query Match 100.0%; Score 106; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
|||||

RESULT 15
ID US-07-881-396A-51 STANDARD; PRT; 26 AA.

XX AC xxxxxx

DT

Sequence 51, Application US/07881396A

Sequence 51, Application US/07881396A

GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/881,396A
FILING DATE: 19920508
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941

INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: N-terminal
SEQUENCE 26 AA; 2844 MW; 3796 CN;

Query Match 100.0%; Score 106; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.03e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 11 HEIKKVLVPGCHGS 24
QY 1 HEIKKVLVPGCHGS 14
|||||

Search completed: Sun Apr 2 17:55:01 2000
Job time : 12 secs.

[illegible]

(TM)

* * *

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search. using Smith-Waterman algorithm

Run on: Sun Apr 2 18:05:25 2000; MasPar time 13.09 Seconds

Tabular output not generated.

Title: >US-09-362-731-4

Description: (1-40) from US09362731, pep

perfect score: 278

Scoring table: PAM 150

Country: **FR: 13**
Gap: **11**

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-pending

1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 24.786; Variance 85.660; scale 0.289

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	Pred. No.
	Score	Match	Length			
1	278	100.0	40	US-09-362-	Sequence 4, Applicatio	4.93e-21
2	110	39.6	347	US-08-443-	Sequence 58, Applicati	2.37e-02
3	110	39.6	347	US-09-004-	Sequence 58, Applicati	2.37e-02
4	110	39.6	347	US-08-054-	Sequence 58, Applicati	2.37e-02
5	109	39.2	23	US-08-944-	Sequence 3, Applicatio	3.00e-02
6	109	39.2	23	US-08-537-	Sequence 25, Applicati	3.00e-02
7	109	39.2	23	US-08-944-	Sequence 3, Applicatio	3.00e-02
8	109	39.2	24	US-08-944-	Sequence 11, Applicati	3.00e-02
9	109	39.2	24	US-08-944-	Sequence 11, Applicati	3.00e-02
10	109	39.2	25	US-09-082-	Sequence 1063, Applic	3.00e-02
11	109	39.2	25	US-08-651-	Sequence 15, Applicati	3.00e-02
12	109	39.2	25	US-08-364-	Sequence 24, Applicati	3.00e-02
13	109	39.2	25	US-08-370-	Sequence 24, Applicati	3.00e-02
14	109	39.2	25	US-09-248-	Sequence 85, Applicati	3.00e-02
15	109	39.2	25	US-09-082-	Sequence 1063, Applic	3.00e-02
16	109	39.2	25	PCT-US99-1	Sequence 1063, Applic	3.00e-02
17	109	39.2	25	PCT-US99-0	Sequence 85, Applicati	3.00e-02
18	109	39.2	25	US-08-487-	Sequence 24, Applicati	3.00e-02
19	109	39.2	25	US-08-481-	Sequence 24, Applicati	3.00e-02
20	109	39.2	25	US-07-801-	Sequence 24, Applicati	3.00e-02

DE Sequence 58, Application US/08443862
XX
CC Sequence 58, Application US/08443862
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/443,862
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:

CC ISSUE:
CC PAGES:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;
Query Match 39.6%; Score 110; DB 8; Length 347;
Best Local Similarity 63.3%; Pred No. 2,37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;
Db 92 QNVKITYGA-CPKYVKNTLKLATGMNV 120
|||::| |||||
QY 6 QNTLKLATGKGGPKYVKNTLKLATGKGV 35
RESULT 3
ID US-09-004-422-58 STANDARD: PRT: 347 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 58, Application US/09004422
XX
CC Sequence 58, Application US/09004422
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
CC STREET: 2033 K Street, N.W., #800
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20006
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/004,422
CC FILING DATE: January 8, 1998
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/443,862
CC FILING DATE: May 22, 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/229,781
CC FILING DATE: April 19, 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/054,016
CC FILING DATE: April 29, 1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-721-8200
CC TELEFAX: 202-721-8250
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide

CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 15; Length 347;
Best Local Similarity 63.3%; Pred. No. 2.37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKQNTKLATGMNV 120
QY 6 QNTLKLATGKKGPKYVKQNTKLATGKKG 35

RESULT 4
ID US-08-054-016B-58 STANDARD; PRT; 347 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 58, Application US/08054016B
XX Sequence 58, Application US/08054016B
CC GENERAL INFORMATION:
CC APPLICANT: Yoshinobu OKUNO et al.
CC TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY
CC NUMBER OF SEQUENCES: 58
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wenderoth, Lind & Ponack
CC STREET: 805 Fifteenth Street, N.W., #700
CC City: Washington
CC STATE: D.C.
CC COUNTRY: U.S.A.
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: MS-DOS

CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/054,016B
CC FILING DATE: April 29, 1993
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warren M. Cheek, Jr.
CC REGISTRATION NUMBER: 33,367
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-371-8850
CC TELEFAX:
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 58:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 347 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FRAGMENT TYPE:
CC ORIGINAL SOURCE:
CC ORGANISM:
CC STRAIN:
CC INDIVIDUAL ISOLATE:
CC DEVELOPMENTAL STAGE:
CC HAPLOTYPE:
CC TISSUE TYPE:
CC CELL LINE:
CC ORGANELLE:
CC IMMEDIATE SOURCE:
CC LIBRARY:
CC CLONE:
CC POSITION IN GENOME:
CC CHROMOSOME/SEGMENT:
CC MAP POSITION:
CC UNITS:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
CC IDENTIFICATION METHOD:
CC OTHER INFORMATION:
CC PUBLICATION INFORMATION:
CC AUTHORS:
CC TITLE:
CC JOURNAL:
CC VOLUME:
CC ISSUE:
CC DATE:
CC DOCUMENT NUMBER:
CC FILING DATE:
CC PUBLICATION DATE:
CC RELEVANT RESIDUES IN SEQ ID NO:
SQ SEQUENCE 347 AA; 39089 MW; 567076 CN;

Query Match 39.6%; Score 110; DB 4; Length 347;
Best Local Similarity 63.3%; Pred. No. 2.37e-02;
Matches 19; Conservative 4; Mismatches 6; Indels 1; Gaps 1;

Db 92 QNVNKITYGA-CPKYVKQNTKLATGMNV 120
QY 6 QNTLKLATGKKGPKYVKQNTKLATGKKG 35

RESULT 5
ID US-08-944-147-3 STANDARD; PRT; 23 AA.

XX Sequence 15, Application US/08651925
DE
XX Sequence 15, Application US/08651925
CC GENERAL INFORMATION:
CC APPLICANT: Langlade-Demoyen, Pierre
CC APPLICANT: Lone, Yu-Chun
CC APPLICANT: Kourilsky, Philippe
CC APPLICANT: Abastado, Jean-Pierre
CC TITLE OF INVENTION: Methods of Using Major
CC TITLE OF INVENTION: Histocompatibility Complex/Peptide Complexes to Obtain or
CC TITLE OF INVENTION: Purify Antigen-Specific T Cells and to Stimulate T Cells
CC NUMBER OF SEQUENCES: 87
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner, L.L.P.
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/651,925
CC FILING DATE: 21-MAY-1997
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 05243.0004-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202) 408-4000
CC TELEFAX: (202) 408-4400
CC INFORMATION FOR SEQ ID NO: 15:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA: 2904 MW; 3266 CN;
SQ
Query Match 39.2%; Score 109; DB 11; Length 25;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 2 PKYVQNTLKLATGMNV 19
QY 18 PKYVQNTLKLATGKGV 35
RESULT 12
ID US-08-364-002-24 STANDARD; PRT; 25 AA.
AC xxxxxx
XX
XX
DE Sequence 24, Application US/08364002
XX Sequence 24, Application US/08364002
CC GENERAL INFORMATION:
CC APPLICANT: Mottez, Estelle
CC APPLICANT: Abastado, Jean-Pierre
CC APPLICANT: Kourilsky, Philippe
CC APPLICANT: Lone, Yu-Chun
CC TITLE OF INVENTION: Altered Major Histocompatibility Complex
CC TITLE OF INVENTION:
CC NUMBER OF SEQUENCES: 127
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/364,002
CC FILING DATE:
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/792,473
CC FILING DATE: 15-NOV-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 03495.0106-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 24:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA: 2904 MW; 3266 CN;
SQ
Query Match 39.2%; Score 109; DB 7; Length 25;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 2 PKYVQNTLKLATGMNV 19
QY 18 PKYVQNTLKLATGKGV 35
RESULT 13
ID US-08-370-476-24 STANDARD; PRT; 25 AA.
AC xxxxxx
XX
XX
DE Sequence 24, Application US/08370476
XX Sequence 24, Application US/08370476
CC GENERAL INFORMATION:
CC APPLICANT: Mottez, Estelle
CC APPLICANT: Abastado, Jean-Pierre
CC APPLICANT: Kourilsky, Philippe
CC APPLICANT: Lone, Yu-Chun
CC APPLICANT: Ojcius, David
CC APPLICANT: Casrouge, Armanda
CC TITLE OF INVENTION: Altered Major Histocompatibility Complex
CC TITLE OF INVENTION:
CC NUMBER OF SEQUENCES: 127
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W., Suite 700
CC CITY: Washington
CC STATE: D.C.
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/370,476
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/117,575
CC FILING DATE: 07-SEP-1993
CC APPLICATION NUMBER: US 08/072,787
CC FILING DATE: 06-JUN-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/801,818
CC FILING DATE: 05-DEC-1991
CC APPLICATION DATA:
CC APPLICATION NUMBER: US 07/792,473
CC FILING DATE: 15-NOV-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 05243.0001-01000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 24:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA; 2904 MW; 3266 CN;

Query Match 39.2%; Score 109; DB 7; Length 25;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 PKYVKQNTLKLATGMNV 19
|||||
QY 18 PKYVKQNTLKLATGKGV 35

RESULT 14
ID US-09-248-588-85 STANDARD; PRT; 25 AA.

AC xxxxxx

XX

XX

DT

XX

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 85, Application US/09248588
Sequence 85, Application US/09248588
GENERAL INFORMATION:
APPLICANT: Birkett, Ashley J.
TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and
TITLE OF INVENTION: their Derivatives
FILE REFERENCE: SYN-101 4564/69529
CURRENT APPLICATION NUMBER: US/09/248,588
EARLIER FILING DATE: 1999-02-11
EARLIER APPLICATION NUMBER: 60/074537
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 85
LENGTH: 25
TYPE: PRT
ORGANISM: Influenza virus
SEQUENCE 25 AA; 2904 MW; 3266 CN;

Query Match 39.2%; Score 109; DB 17; Length 25;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 PKYVKQNTLKLATGMNV 19
|||||

QY 18 PKYVKQNTLKLATGKGV 35

RESULT 15

ID US-09-082-279-1063 STANDARD; PRT; 25 AA.

AC xxxxxx

XX

XX

DT

XX

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 39.2%; Score 109; DB 15; Length 25;
Best Local Similarity 83.3%; Pred. No. 3.00e-02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 2 PKYVKQNTLKLATGMNV 19
|||||

QY 18 PKYVKQNTLKLATGKGV 35

Search completed: Sun Apr 2 18:05:42 2000
Job time : 17 secs.

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Apr 2 18:10:34 2000; MasPar time 12.24 Seconds
278.714 Million cell updates/sec
Tabular output not generated.

Title: >US-09-362-731-5
Description: (1-32) from US09362731.pep
Perfect Score: 241
Sequence: 1 QV1KANSKFIGITELGGCHGSEPCNIHRGKPF 32

Scoring table: PAM 150
Gap 11

Searched: 680877 seqs, 106649801 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1:PG 2:U60 3:07 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:U94 20:NEWP 21:NEWU60 22:NEWU6 23:NEWU8
24:NEWU9

Statistics: Mean 24.794; Variance 65.012; scale 0.381

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	241	100.0	32	18	US-09-362-Sequence 5, Applicatio	3.35e-23
2	221	91.7	137	18	US-09-362-Sequence 3, Applicatio	3.33e-20
3	122	50.6	25	8	US-08-445-Sequence 104, Applicat	5.99e-06
4	122	50.6	25	9	US-08-484-Sequence 42, Applicati	5.99e-06
5	122	50.6	25	3	US-07-881-Sequence 104, Applicat	5.99e-06
6	122	50.6	25	3	US-07-881-Sequence 42, Applicat	5.99e-06
7	122	50.6	25	8	US-08-445-Sequence 42, Applicati	5.99e-06
8	122	50.6	25	6	US-08-227-Sequence 42, Applicati	5.99e-06
9	122	50.6	25	9	US-08-484-Sequence 104, Applicat	5.99e-06
10	122	50.6	35	9	US-08-460-Sequence 7, Applicatio	5.99e-06
11	122	50.6	61	9	US-08-484-Sequence 162, Applicat	5.99e-06
12	122	50.6	61	6	US-08-227-Sequence 162, Applicat	5.99e-06
13	122	50.6	61	8	US-08-445-Sequence 162, Applicat	5.99e-06
14	122	50.6	84	6	US-08-227-Sequence 161, Applicat	5.99e-06
15	122	50.6	84	8	US-08-445-Sequence 161, Applicat	5.99e-06
16	122	50.6	84	9	US-08-484-Sequence 161, Applicat	5.99e-06
17	122	50.6	92	6	US-08-227-Sequence 160, Applicat	5.99e-06
18	122	50.6	92	8	US-08-445-Sequence 160, Applicat	5.99e-06
19	122	50.6	92	9	US-08-484-Sequence 160, Applicat	5.99e-06
20	122	50.6	129	14	US-08-930-Sequence 8, Applicatio	5.99e-06

21	122	50.6	129	6	US-08-227-Sequence 157, Applicat	5.99e-06
22	122	50.6	129	9	US-08-484-Sequence 157, Applicat	5.99e-06
23	122	50.6	129	9	US-08-484-Sequence 159, Applicat	5.99e-06
24	122	50.6	129	8	US-08-445-Sequence 159, Applicat	5.99e-06
25	122	50.6	129	14	US-08-930-Sequence 6, Applicatio	5.99e-06
26	122	50.6	129	14	US-08-930-Sequence 24, Applicati	5.99e-06
27	122	50.6	129	14	US-08-930-Sequence 18, Applicati	5.99e-06
28	122	50.6	129	14	US-08-930-Sequence 10, Applicati	5.99e-06
29	122	50.6	129	14	US-08-930-Sequence 2, Applicatio	5.99e-06
30	122	50.6	129	14	US-08-930-Sequence 12, Applicati	5.99e-06
31	122	50.6	129	14	US-08-930-Sequence 14, Applicati	5.99e-06
32	122	50.6	129	6	US-08-227-Sequence 8, Applicatio	5.99e-06
33	122	50.6	129	3	US-07-881-Sequence 8, Applicatio	5.99e-06
34	122	50.6	129	8	US-08-445-Sequence 8, Applicatio	5.99e-06
35	122	50.6	129	8	US-08-445-Sequence 157, Applicat	5.99e-06
36	122	50.6	129	9	US-08-484-Sequence 8, Applicatio	5.99e-06
37	122	50.6	130	6	US-08-227-Sequence 158, Applicat	5.99e-06
38	122	50.6	130	9	US-08-484-Sequence 158, Applicat	5.99e-06
39	122	50.6	130	8	US-08-445-Sequence 158, Applicat	5.99e-06
40	122	50.6	142	9	US-08-483-Sequence 2, Applicatio	5.99e-06
41	122	50.6	142	9	US-08-483-Sequence 4, Applicatio	5.99e-06
42	122	50.6	145	9	US-08-460-Sequence 6, Applicatio	5.99e-06
43	122	50.6	146	9	US-08-484-Sequence 4, Applicatio	5.99e-06
44	122	50.6	146	8	US-08-445-Sequence 4, Applicatio	5.99e-06
45	122	50.6	146	3	US-07-881-Sequence 4, Applicatio	5.99e-06

ALIGNMENTS

RESULT 1
ID US-09-362-731-5 STANDARD; PRT: 32 AA.

AC xxxxxx

DT

DE

Sequence 5, Application US/09362731

XX

Sequence 5, Application US/09362731

CC

GENERAL INFORMATION:

CC

APPLICANT: U.C.B. S.A.

CC

TITLE OF INVENTION: COMPOUND AND METHOD FOR THE PREVENTION AND/OR THE

CC

FILE REFERENCE: P. UCB.09/WO

CC

CURRENT APPLICATION NUMBER: US/09/362,731

CC

CURRENT FILING DATE: 1999-07-29

CC

NUMBER OF SEQ ID NOS: 17

CC

SOFTWARE: PatentIn Ver. 2.1

CC

SEQ ID NO 5

CC

LENGTH: 32

CC

TYPE: PRT

CC

ORGANISM: Artificial Sequence

CC

FEATURE:

CC

OTHER INFORMATION: Description of Artificial Sequence: synthetic

CC

OTHER INFORMATION: peptide

SQ

SEQUENCE 32 AA; 3503 MW; 5050 CN;

Query Match 100.0%; Score 241; DB 18; Length 32;

Best Local Similarity 100.0%; Pred. No. 3.35e-23;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QV1KANSKFIGITELGGCHGSEPCNIHRGKPF 32

|||||

QY 1 QV1KANSKFIGITELGGCHGSEPCNIHRGKPF 32

RESULT 2

ID US-09-362-731-3 STANDARD; PRT: 137 AA.

XX xxxxxx

XX

XX

XX

CC		TELEPHONE:	(617) 466-6000
CC		TELEFAX:	(617) 466-6040
CC		INFORMATION FOR SEQ ID NO:	104:
CC		SEQUENCE CHARACTERISTICS:	
CC		LENGTH:	25 amino acids
CC		TYPE:	amino acid
CC		TOPOLOGY:	linear
CC		MOLECULE TYPE:	peptide
CC		FRAGMENT TYPE:	N-terminal
CC		SEQUENCE	25 AA; 2781 MW; 3272 CN;
SQ			
		Query Match	50.6%; Score 122; DB 8; Length 25;
		Best Local Similarity	77.9%; Pred. No. 5.99e-06;
		Matches	14; Conservative 2; Mismatches 2; Indels 0;
Dd	8	V D G C H G S D P C I I H R G K P F	25
Qy	15	L G G C H G S E P C N I I H R G A P F	32
		: ::	
RESULT	4		
ID	US-08-484-296-42	STANDARD;	PRT; 25 AA.
XX			
AC	xxxxxx		
DT			
XX			
XX			
DE			
XX			
XX			
Sequence 42,	Application US/08484296		
Sequence 42,	Application US/08484296		
GENERAL INFORMATION:			
APPLICANT:	Garman, Richard		
APPLICANT:	Greenstein, Julia		
APPLICANT:	Kuo, Mei-chang		
APPLICANT:	Rogers, Bruce		
APPLICANT:	Franzen, Henry		
APPLICANT:	Chen, Xian		
APPLICANT:	Evans, Sean		
APPLICANT:	Shaked, Ze'ev		
TITLE OF INVENTION:	T CELL EPITOPES OF THE MAJOR ALLERGENS		
TITLE OF INVENTION:	FROM DERMATOPHAGOIDES (HOUSE DUST MITE)		
NUMBER OF SEQUENCES:	207		
CORRESPONDENCE ADDRESS:			
ADDRESSES:	IMMULOGIC PHARMACEUTICAL CORPORATION		
STREET:	610 LINCOLN STREET		
CITY:	WALTHAM		
STATE:	MA		
COUNTRY:	USA		
ZIP:	02154		
COMPUTER READABLE FORM:			
MEDIUM TYPE:	Floppy disk		
COMPUTER:	IBM PC compatible		
OPERATING SYSTEM:	PC-DOS/MS-DOS		
SOFTWARE:	ASCII TEXT		
CURRENT APPLICATION DATA:			
APPLICATION NUMBER:	US/08/484,296		
FILING DATE:			
CLASSIFICATION:	435		
PRIOR APPLICATION DATA:			
APPLICATION NUMBER:	08/445,307		
FILING DATE:	07 June 1995		
ATTORNEY/AGENT INFORMATION:			
NAME:	CRAIG, ANNE I.		
REGISTRATION NUMBER:	32,976		
REFERENCE/DOCKET NUMBER:	017,605		
TELECOMMUNICATION INFORMATION:			
TELEPHONE:	(617) 466-6000		
TELEFAX:	(617) 466-6040		
INFORMATION FOR SEQ ID NO:	42:		
SEQUENCE CHARACTERISTICS:			
LENGTH:	25 amino acids		
TYPE:	amino acid		
TOPOLOGY:	linear		

```
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
SQ SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match
Best Local Similarity 50.6%; Score 122; DB 9; Length 25;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 10 GCHGSEPCIIHRGKPF 25
QY 17 LGCHGSEPCNIHRGKPF 32

RESULT 5
ID US-07-881-396A-104 STANDARD; PRT; 25 AA.
XX
AC
XX
DT
XX
DE
XX
XX
CC Sequence 104, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-5941
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 104:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
SQ SEQUENCE 25 AA; 2781 MW; 3272 CN;

Query Match
Best Local Similarity 50.6%; Score 122; DB 3; Length 25;
Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 8 VDGCHGSDPCIIHRGKPF 25
QY 15 LGCHGSEPCNIHRGKPF 32

RESULT 6
ID US-07-881-396A-42 STANDARD; PRT; 25 AA.
XX
AC
XX
DT
XX
DE
XX
XX
CC Sequence 104, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-5941
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 104:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
SQ SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match
Best Local Similarity 50.6%; Score 122; DB 3; Length 25;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 10 GCHGSEPCIIHRGKPF 25
QY 17 LGCHGSEPCNIHRGKPF 32

RESULT 7
ID US-08-445-307A-42 STANDARD; PRT; 25 AA.
XX
AC
XX
DT
XX
DE
XX
XX
CC Sequence 42, Application US/08445307A
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
```

```
XX
AC
XX
DT
XX
DE
XX
XX
CC Sequence 42, Application US/07881396A
CC GENERAL INFORMATION:
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 119
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/881,396A
CC FILING DATE: 19920508
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-5941
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
SQ SEQUENCE 25 AA; 2782 MW; 3408 CN;

Query Match
Best Local Similarity 50.6%; Score 122; DB 3; Length 25;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 10 GCHGSEPCIIHRGKPF 25
QY 17 LGCHGSEPCNIHRGKPF 32

RESULT 7
ID US-08-445-307A-42 STANDARD; PRT; 25 AA.
XX
AC
XX
DT
XX
DE
XX
XX
CC Sequence 42, Application US/08445307A
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
```

CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
Query Match 50.6%; Score 122; DB 8; Length 25;
Best Local Similarity 93.8%; Pred. No. 5,99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 10 GCHGSEPCIIHRGKPF 25
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 8
ID US-08-227-772-42 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 42, Application US/08227772
Sequence 42, Application US/08227772
GENERAL INFORMATION:
APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 42:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2782 MW; 3408 CN;
Query Match 50.6%; Score 122; DB 6; Length 25;
Best Local Similarity 93.8%; Pred. No. 5,99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 10 GCHGSEPCIIHRGKPF 25
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 9
ID US-08-484-296-104 STANDARD; PRT; 25 AA.
XX
AC xxxxxx
XX
DT
XX
Sequence 104, Application US/08484296
Sequence 104, Application US/08484296
GENERAL INFORMATION:
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
APPLICANT: Kuo, Mei-chang
APPLICANT: Rogers, Bruce
APPLICANT: Franzen, Henry
APPLICANT: Chen, Xian
APPLICANT: Evans, Sean
APPLICANT: Shaked, Ze'ev
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
STREET: 610 LINCOLN STREET
CITY: WALTHAM
STATE: MA
COUNTRY: USA
ZIP: 02154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,296
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 104:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 25 AA; 2781 MW; 3272 CN;
CC
CC Query Match 50.6%; Score 122; DB 9; Length 25;
CC Best Local Similarity 77.8%; Pred. No. 5.99e-06;
CC Matches 14; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
CC Db 8 VDGCHGSDPCIIHRGKPF 25
CC : |||||:|||||
CC QY 15 LGGCHGSEPCNIHRGKPF 32
CC
CC RESULT 10
CC ID US-08-460-040-7 STANDARD; PRT; 35 AA.
CC AC xxxxxx
CC XX
CC DT
CC DE
CC XX
CC DE
CC XX
CC Sequence 7, Application US/08460040
CC
CC Sequence 7, Application US/08460040
CC GENERAL INFORMATION:
CC APPLICANT: Thomas, Wayne R.
CC TITLE OF INVENTION: Cloning of Mite Allergens
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent in Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/460,040
CC FILING DATE: 2-JUNE-95
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/162,722
CC FILING DATE: 8-NOV-93
CC APPLICATION NUMBER: 07/458,642
CC FILING DATE: 13-FEB-90
CC APPLICATION NUMBER: PCT/AU88/00195
CC FILING DATE: 17-JUNE-88
CC APPLICATION NUMBER: PI 2523/87
CC FILING DATE: 18-JUNE-87
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Amy E. Mandragouras
CC REGISTRATION NUMBER: 36,207
CC REFERENCE/DOCKET NUMBER: IMI-021CN2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941

CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 35 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 35 AA; 3870 MW; 6469 CN;
CC
CC Query Match 50.6%; Score 122; DB 9; Length 35;
CC Best Local Similarity 93.8%; Pred. No. 5.99e-06;
CC Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CC
CC Db 20 GCHGSEPCIIHRGKPF 35
CC : |||||:|||||
CC QY 17 GCHGSEPCNIHRGKPF 32
CC
CC RESULT 11
CC ID US-08-484-296-162 STANDARD; PRT; 61 AA.
CC AC xxxxxx
CC XX
CC DT
CC DE
CC XX
CC DE
CC XX
CC Sequence 162, Application US/08484296
CC
CC Sequence 162, Application US/08484296
CC GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/484,296
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/445,307
CC FILING DATE: 07 June 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017.605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 162:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 61 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal

US-08-445-307A-162	STANDARD;	PRT: 61 AA.
xxxxxx		
Sequence 162, Application US/08445307A		
Sequence 162, Application US/08445307A		
GENERAL INFORMATION:		
APPLICANT: Garman, Richard		
APPLICANT: Greenstein, Julia		
APPLICANT: Kuo, Mei-chang		
APPLICANT: Rogers, Bruce		
APPLICANT: Franzen, Henry		
APPLICANT: Chen, Xian		
APPLICANT: Evans, Sean		
APPLICANT: Shaked, Ze'ev		
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS		
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)		
NUMBER OF SEQUENCES: 207		
CORRESPONDENCE ADDRESS:		
ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION		
STREET: 610 LINCOLN STREET		
CITY: WALTHAM		
STATE: MA		
COUNTRY: USA		
ZIP: 02154		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
COMPUTER: IBM PC Compatible		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: ASCII TEXT		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/445.307A		
FILING DATE: 07 June 1995		
CLASSIFICATION: 435		
ATTORNEY/AGENT INFORMATION:		
NAME: CRAIG, ANNE I.		
REGISTRATION NUMBER: 32,976		
REFERENCE/DOCKET NUMBER: 017.6US		
TELECOMMUNICATION INFORMATION:		
TELEPHONE: (617) 466-6000		
TELEFAX: (617) 466-6040		
INFORMATION FOR SEQ ID NO: 162:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 61 amino acids		
TYPE: amino acid		
TOPOLOGY: linear		
MOLECULE TYPE: peptide		
FRAGMENT TYPE: internal		
SEQUENCE 61 AA; 6669 MW; 17328 CN;		
Query Match	50.6%;	Score 122; DB 8; Length 61;
Best Local Similarity	93.8%;	Pred. No. 5.99e-06;
Matches	15; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Db	20 GCHGSEPCIIHRGKPF 35	
Qy	17 GCHGSEPCNIHRGKPF 32	
RESULT 14		
ID US-08-227-772-161	STANDARD;	PRT: 84 AA.
XX		
AC	xxxxxx	
XX		
DT		
XX		
DE	Sequence 161, Application US/082227772	
XX		
CC	Sequence 161, Application US/082227772	
CC	GENERAL INFORMATION:	

CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 197
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 STATE STREET SUITE 510
CC CITY: BOSTON
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/227,772
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/881,396
CC FILING DATE: 08-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: MANDRAGOURAS, AMY E.
CC REGISTRATION NUMBER: P36,207
CC REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 227-7400
CC TELEFAX: (617) 227-5941
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;
SQ
Query Match 50.68; Score 122; DB 6; Length 84;
Best Local Similarity 93.8%; Pred. No. 5.99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
RESULT 15
ID US-08-445-307A-161 STANDARD; PRT; 84 AA.
AC xxxxxx
XX
XX
XX
XX
XX
XX
XX
XX
XX
XX
Sequence 161, Application US/08445307A
Sequence 161, Application US/08445307A
GENERAL INFORMATION:
CC APPLICANT: Garman, Richard
CC APPLICANT: Greenstein, Julia
CC APPLICANT: Kuo, Mei-chang
CC APPLICANT: Rogers, Bruce
CC APPLICANT: Franzen, Henry
CC APPLICANT: Chen, Xian
CC APPLICANT: Evans, Sean
CC APPLICANT: Shaked, Ze'ev
CC TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
CC TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
CC STREET: 610 LINCOLN STREET
CC CITY: WALTHAM
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII TEXT
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/445,307A
CC FILING DATE: 07 June 1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CRAIG, ANNE I.
CC REGISTRATION NUMBER: 32,976
CC REFERENCE/DOCKET NUMBER: 017,605
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 466-6000
CC TELEFAX: (617) 466-6040
CC INFORMATION FOR SEQ ID NO: 161:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 84 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 84 AA; 9150 MW; 34277 CN;
SQ
Query Match 50.6%; Score 122; DB 8; Length 84;
Best Local Similarity 93.8%; Pred. No. 5.99e-06;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db 20 GCHGSEPCIIHRGKPF 35
QY 17 GCHGSEPCNIHRGKPF 32
Search completed: Sun Apr 2 18:10:48 2000
Job time : 14 secs.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 8, 2000, 01:04:05 ; Search time 624.95 Seconds
(without alignments)
-1008.099 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagccgttctaagcgccgcg 175

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5086966 seqs, -1800032204 residues

Total number of hits satisfying chosen parameters: 10173922

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending_Patents_NA:*

- 1: /cgn2_6/ptodata/2/pna/PTUS9_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US081A_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US081B_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US081C_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US081D_COMB.seq:*
- 8: /cgn2_6/ptodata/2/pna/US082A_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US082B_COMB.seq:*
- 10: /cgn2_6/ptodata/2/pna/US082C_COMB.seq:*
- 11: /cgn2_6/ptodata/2/pna/US083A_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US083B_COMB.seq:*
- 13: /cgn2_6/ptodata/2/pna/US084A_COMB.seq:*
- 14: /cgn2_6/ptodata/2/pna/US084B_COMB.seq:*
- 15: /cgn2_6/ptodata/2/pna/US084C_COMB.seq:*
- 16: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
- 17: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
- 18: /cgn2_6/ptodata/2/pna/US087A_COMB.seq:*
- 19: /cgn2_6/ptodata/2/pna/US087B_COMB.seq:*
- 20: /cgn2_6/ptodata/2/pna/US087C_COMB.seq:*
- 21: /cgn2_6/ptodata/2/pna/US088A_COMB.seq:*
- 22: /cgn2_6/ptodata/2/pna/US088B_COMB.seq:*
- 23: /cgn2_6/ptodata/2/pna/US088C_COMB.seq:*
- 24: /cgn2_6/ptodata/2/pna/US088D_COMB.seq:*
- 25: /cgn2_6/ptodata/2/pna/US089A_COMB.seq:*
- 26: /cgn2_6/ptodata/2/pna/US089B_COMB.seq:*
- 27: /cgn2_6/ptodata/2/pna/US089C_COMB.seq:*
- 28: /cgn2_6/ptodata/2/pna/US089D_COMB.seq:*
- 29: /cgn2_6/ptodata/2/pna/US089E_COMB.seq:*
- 30: /cgn2_6/ptodata/2/pna/US089F_COMB.seq:*
- 31: /cgn2_6/ptodata/2/pna/US089G_COMB.seq:*
- 32: /cgn2_6/ptodata/2/pna/US089H_COMB.seq:*
- 33: /cgn2_6/ptodata/2/pna/US089I_COMB.seq:*
- 34: /cgn2_6/ptodata/2/pna/US090A_COMB.seq:*
- 35: /cgn2_6/ptodata/2/pna/US090B_COMB.seq:*
- 36: /cgn2_6/ptodata/2/pna/US090C_COMB.seq:*
- 37: /cgn2_6/ptodata/2/pna/US090D_COMB.seq:*
- 38: /cgn2_6/ptodata/2/pna/US090E_COMB.seq:*
- 39: /cgn2_6/ptodata/2/pna/US091A_COMB.seq:*
- 40: /cgn2_6/ptodata/2/pna/US091B_COMB.seq:*
- 41: /cgn2_6/ptodata/2/pna/US091C_COMB.seq:*
- 42: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
- 43: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
- 44: /cgn2_6/ptodata/2/pna/US092C_COMB.seq:*

Result No.	Score	Query Match %	ID	Description
1	175	100.0	US-09-362-731-6	Sequence 6, Appl
2	101	57.7	US-09-362-731-12	Sequence 12, Appl
3	100	57.1	US-09-362-731-11	Sequence 11, Appl
4	93	53.1	US-09-362-731-9	Sequence 9, Appl
5	48.2	27.5	US-07-881-396A-3	Sequence 3, Appl
6	48.2	27.5	US-08-227-772-3	Sequence 3, Appl
7	48.2	27.5	US-08-445-307A-3	Sequence 3, Appl
8	48.2	27.5	US-08-484-296-3	Sequence 3, Appl
9	48.2	27.5	US-08-460-040-5	Sequence 5, Appl
10	43.4	24.8	US-08-930-264-15	Sequence 15, Appl
11	42.6	24.3	US-08-930-264-19	Sequence 19, Appl
12	41.8	23.9	US-08-930-264-1	Sequence 1, Appl
13	41.8	23.9	US-08-930-264-3	Sequence 3, Appl
14	41.8	23.9	US-08-930-264-5	Sequence 5, Appl

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES


```

; LOCATION: 69..509
US-07-881-396A-3

Query Match      27.5%; Score 48.2; DB 2; Length 588;
Best Local Similarity 94.3%; Pred. No. 5.2e-06;
Matches 50; Conservative 0; Mismatches 3; Indels 0;

QY 61 taggaggttccatcggttcagaaccatgtatcatctcgttggtaaaccattc 113
|||
Db 172 TACCAGATGGCATGGTTCAGAACCATGTATCATTCATCGTGGTAAACCATTC 224
|||||

```

RESULT 6
US-08-227-772-3
; Sequence 3, Application us/08227772
; GENERAL INFORMATION:
; APPLICANT: Kuo, Mei-chang
; APPLICANT: Garman, Richard
; APPLICANT: Greenstein, Julia
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
; TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

APPLICANT: Kuo, Mei-chang
APPLICANT: Garman, Richard
APPLICANT: Greenstein, Julia
TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
TITLE OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/227,772
FILING DATE: 14-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/881,396
FILING DATE: 08-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IMI-012/IPC-017C
TELECOMMUNICATION INFORMATION:

```

, TELEPHONE: (617) 227-7400
,
, TELEFAX: (617) 227-5941
,
, INFORMATION FOR SEQ ID NO: 3:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 588 base pairs
,
, TYPE: nucleic acid
,
, STRANDEDNESS: single
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: cDNA
,
, FEATURE:
,
, NAME/KEY: CDS
,
, LOCATION: 65..509
,
, US-08-227-772-3

```

```

;
; NAME/KEY: CDS
; LOCATION: 65..509
US-08-227-772-3

Query Match          27.5%; Score 48.2; DB 8; Length 588;
Best Local Similarity 94.3%; Pred. No. 5.2e-06;
Matches 50; Conservative 0; Mismatches 3; Indels 0;

Qy 61 taggagggtgcgatggttcagaaacatgtatcattcatcggtgtaaacattc 113
    |||
Db 172 TACCAGGATGCCATGGTTCAGAACCAATGATCATTCATCGTGGTAACCAATC 224

RESULT 7
US-08-445-307A-3

```

Query Match	27.5%	Score 48.2;	DB 15;	Length 588;
Best Local Similarity	94.3%;	Pred. No. 5.2e-06;		
Matches 50;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

Qy	61	taggaggttgccatggttcagaaccattgatcatctcgtagaacattc	113
Db	172	TACCAAGGATGCCATGGTTCAGAACCATTGTATCATCTCGTTAAACCATTC	224

```

RESULT          9
US-08-460-040-5
; Sequence 5, Application US/08460040
; GENERAL INFORMATION:
;   APPLICANT: Thomas, Wayne R.
;   TITLE OF INVENTION: Cloning of Mite Allergens
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: LAHIVE & COCKFIELD
;   STREET: 60 State Street, suite 510
;   CITY: Boston
;   STATE: Massachusetts
;   COUNTRY: USA
;   ZIP: 02109-1875

```

```

1 ZIP: 02105-1873
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/08/460,040
16
17 FILING DATE: 2-JUNE-95
18
19 CLASSIFICATION: 435
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: 08/162,722

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

RESULT 11
US-08-930-264-19
; Sequence 19, Application US/08930264
; GENERAL INFORMATION:
; APPLICANT: NISHIYAMA, Chiharu
; APPLICANT: YUKI, Yoshifumi
; APPLICANT: OKUMURA, Yasushi
; TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
; TITLE OF INVENTION: PRODUCTION THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,264
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-93236
; FILING DATE: 28-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00791
; FILING DATE: 27-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NISHIYAMA-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..387
US-08-930-264-19

Query Match 24.3%; Score 42.6; DB 27; Length 390;
Best Local Similarity 91.8%; Pred. No. 0.00035;
Matches 45; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 65 aggttcgcatggttcagaccatgtatcatcattcggtgtaaacattc 113
|||||
Db 57 AGGTTCGTCATGTCGATCCATCATCCATCCATCGTGGTAAACCATTC 105

RESULT 12
US-08-930-264-1
; Sequence 1, Application US/08930264
; GENERAL INFORMATION:
; APPLICANT: NISHIYAMA, Chiharu
; APPLICANT: YUUKI, Toshifumi
; APPLICANT: OKUMURA, Yasushi
; TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
; TITLE OF INVENTION: PRODUCTION THEREOF
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,264
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-93236
; FILING DATE: 28-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/00791
; FILING DATE: 27-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: NISHIYAMA-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 390 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..387
US-08-930-264-1

Query Match 23.9%; Score 41.8; DB 27; Length 390;
Best Local Similarity 86.8%; Pred. No. 0.00066;
Matches 46; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 61 taggaggttcgcatggttcagaccatgtatcatcattcggtgtaaacattc 113
|||||
Db 53 TCGATGTTGCCATGTTCTGATCCATCATCCATCCATCGTGGTAAACCATTC 105

RESULT 14
US-08-930-264-5
; Sequence 5, Application US/08930264
; GENERAL INFORMATION:
; APPLICANT: NISHIYAMA, Chiharu
; APPLICANT: YUUKI, Toshifumi
; APPLICANT: OKUMURA, Yasushi
; TITLE OF INVENTION: MODIFIED MITE ALLERGEN AND THE
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 7, 2000, 23:43:21 ; Search time 37.87 seconds
(without alignments)
553.249 Million cell updates/sec

Title: US-09-362-731-6
Perfect score: 175
Sequence: 1 gaattccaccatgatcag.....aagccgttctaagcggcgc 175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 214294 seqs, 59861574 residues

Total number of hits satisfying chosen parameters: 428588

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCUS9_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.2	27.5	588	1	US-07-945-288-3
2	48.2	27.5	588	1	US-08-462-831-3
3	48.2	27.5	588	2	US-08-461-809-3
4	48.2	27.5	588	2	US-08-461-441-3
5	48.2	27.5	588	2	US-08-482-142-3
6	48.2	27.5	588	4	US-08-478-572-3
7	48.2	27.5	588	6	PCUS93-08518-3
8	41.8	23.9	513	2	US-08-288-888-3
9	41.8	23.9	513	2	US-08-910-075-3
10	41.8	23.9	513	4	US-08-905-801A-3
11	41.8	23.9	516	2	US-08-288-888-1
12	41.8	23.9	517	3	US-08-910-075-1
13	41.8	23.9	517	4	US-08-905-801A-1
14	40.2	23.0	491	1	US-07-945-288-7
15	40.2	23.0	491	1	US-08-462-831-7
16	40.2	23.0	491	2	US-08-461-809-7
17	40.2	23.0	491	2	US-08-461-441-7
18	40.2	23.0	491	2	US-08-482-142-7
19	40.2	23.0	491	4	US-08-478-572-7
20	40.2	23.0	491	6	PCUS93-08518-7
21	29.2	16.7	1474	2	US-08-465-980-1
22	29.2	16.7	1474	4	US-09-053-303-1
23	29.2	16.7	1474	6	PCUS95-07093-1
24	28.2	16.1	1151	1	US-08-236-754-3
25	28.2	16.1	3055	1	US-08-236-754-1
26	27	15.4	2380	2	US-08-572-951-3
27	27	15.4	3142	4	US-08-956-242-3

C 28 26.8 15.3 5359 4 US-09-070-060-1 Sequence 1, Appli
C 29 26.6 15.2 1315 3 US-08-578-592-4 Sequence 4, Appli
30 2973 2 US-08-451-715A-7 Sequence 7, Appli
31 26.4 15.1 1626 2 US-08-604-333-1 Sequence 1, Appli
32 26.2 15.0 2951 1 US-08-413-118-104 Sequence 104, App
33 26.2 15.0 3000 1 US-08-220-151-1 Sequence 1, Appli
34 26.2 15.0 3000 1 US-08-413-118-1 Sequence 1, Appli
35 26 14.9 1503 1 US-08-307-499-41 Sequence 41, Appli
C 36 26 14.9 14176 1 US-08-307-499-14 Sequence 14, Appli
37 26 14.9 14176 1 US-08-661-052-7 Sequence 7, Appli
38 25.6 14.6 54 3 US-08-551-437-4 Sequence 4, Appli
C 39 25.2 14.4 1734 2 US-08-602-010A-1 Sequence 1, Appli
40 25.2 14.4 5495 2 US-08-680-726A-1 Sequence 2, Appli
C 41 25.2 14.4 5495 2 US-08-680-726A-2 Sequence 2, Appli
42 25.2 14.4 5495 2 US-08-158-232-50 Sequence 50, Appli
C 43 25.2 14.4 2172 1 US-08-611-928-50 Sequence 50, Appli
44 25 14.3 2172 2 US-08-611-928-50 Sequence 50, Appli
45 25 14.3 2172 2 US-08-611-928-50 Sequence 50, Appli

ALIGNMENTS

RESULT 1
US-07-945-288-3
; Sequence 3, Application US/07945288
; Patent No. 5433948
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kew-Yan
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII TEXT
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 580,655
FILING DATE: 11 SEPTEMBER 1990
APPLICATION NUMBER: 458,642
FILING DATE: 13 FEBRUARY 1990
ATTORNEY/AGENT INFORMATION:
NAME: MANDRAGOURAS, AMY E.
REGISTRATION NUMBER: P36,207
REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 69..509
US-07-945-288-3


```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: ASCII TEXT
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/482,142
? FILING DATE: 07-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/445,307
? FILING DATE: 07 June 1995
? ATTORNEY/AGENT INFORMATION:
? NAME: CRAIG, ANNE I.
? REGISTRATION NUMBER: 32,976
? REFERENCE/DOCKET NUMBER: 017.6US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617) 466-6000
? TELEFAX: (617) 466-6040
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 588 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 69..509
? US-08-482-142-3
?
? Query Match 27.5% Score 48.2; DB 2; Length 588;
? Best Local Similarity 94.3%; Pred. No. 2.5e-07;
? Matches 50; Conservative 0; Mismatches 3; Indels 0;
?
? Qy 61 taggaggttgcattggttcagaaacattgatcattcatctgtgtaaacattc 113
? |||
? Db 172 TACCAGAGTGCATGGTTTCAGAACCATTCATTCATTCGTGGTAAACCATTC 224

```

RESULT 6
US-08-478-572-3
; Sequence 3, Application US/08478572
; Patent No. 5968526
; GENERAL INFORMATION:
; APPLICANT: Garman, Richard
; APPLICANT: Greenstein, Julia
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Rogers, Bruce
; APPLICANT: Franzen, Henry
; APPLICANT: Chen, Xian
; APPLICANT: Evans, Sean
; APPLICANT: Shaked, Ze'ev
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS
; OF INVENTION: FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMUNOLOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: MA
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IEM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,572
; FILING DATE: 07-June-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/445,307
; FILING DATE:

9 RESULT


```
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..390
US-07-945-288-7

Query Match      23.0%; Score 40.2; DB 1; Length 491;
Best Local Similarity 84.9%; Pred. No. 0.00011;
Matches 45; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY   61 tagagagttccactgggttcagaaccatgatatcattcatctggtgtaaacattc 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Dd   53 TCGATGGTCCCATGGTTCTGTCATCATTGCATAATCCATCGTGTAACCAATTC 105

RESULT 15
US-08-462-831-7
; Sequence 7, Application US/08462831
; Patent No. 5552142
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,831
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (INT-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 491 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..390
US-08-462-831-7

Query Match      23.0%; Score 40.2; DB 1; Length 491;
Best Local Similarity 84.9%; Pred. No. 0.00011;
Matches 45; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY   61 tagagagttccactgggttcagaaccatgatatcattcatctggtgtaaacattc 113
```

Db 53 TCGATGGTGGCCATGGTTCTGTGATCCATGCATATCCATCCATCGTCGTTAAACCATTC 105

Search completed: April 8, 2000, 02:55:17
Job time: 11516 sec